STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 131743

TO: Ginny Portner
Location: REM-3B02/3C18
Art Unit: 1645
Friday, September 10, 2004

Case Serial Number: 10/729039

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-A-62
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Portner,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
Remsen 1-A-62
Ext. 22524

This Page Blank (uspto)

131743

STIC-Biotech/ChemLib

From: STIC-ILL
Sent: Tuesday, September 07, 2004 8:48 AM
To: STIC-Biotech/ChemLib
Subject: FW: 10/729,039

-----Original Message-----
From: Portner, Ginny
Sent: Tuesday, September 07, 2004 8:48 AM
To: STIC-ILL
Subject: 10/729,039

Please search SEQ ID No 44, 46, 60, 62, 68,50, 52, 73, 77 and 79 (oligo and full sequence please) thanks!

Ginny Portner
Remsen Building
Art Unit 1645
Room E03, B02
(571) 272-0862

****
STAFF USE ONLY

Searcher:_____
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed:_____
Searcher Prep/Rev. Time:___
Online Time:_____

****
Type of Search
NA Sequence: #_____
AA Sequence :#_____
Structure: #_____
Bibliographic:_____
Litigation:_____
Patent Family:_____
Other:_____

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other(Specify):_____

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 27.6553 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	472	100.0	472	2	AAW68393	Aaw68393 Clostridi
2	70	14.8	440	3	AAY77135	Aay77135 Synthetic
3	70	14.8	440	4	AAB04091	Aab04091 Botulism
4	70	14.8	472	2	AAW68394	Aaw68394 Clostridi
5	70	14.8	1070	3	AAY93308	Aay93308 A mangane
6	70	14.8	1095	3	AAY93311	Aay93311 A mangane
7	70	14.8	1169	2	AAW56017	Aaw56017 Recombina
8	70	14.8	1291	2	AAW68392	Aaw68392 Clostridi
9	42	8.9	848	4	AAB04082	Aab04082 Botulism
10	39	8.3	439	4	AAB04085	Aab04085 Botulism
11	31	6.6	473	2	AAW68400	Aaw68400 Clostridi
12	26	5.5	448	2	AAW68399	Aaw68399 Clostridi
13	26	5.5	451	2	AAW68395	Aaw68395 Clostridi
14	26	5.5	451	2	AAW68398	Aaw68398 Clostridi
15	26	5.5	452	2	AAW68396	Aaw68396 Clostridi
16	26	5.5	462	2	AAR95009	Aar95009 Type A ne
17	26	5.5	462	2	AAW68390	Aaw68390 Clostridi
18	26	5.5	462	2	AAW68397	Aaw68397 Clostridi
19	23	4.9	393	4	AAG68149	Aag68149 Streptomy
20	23	4.9	393	4	AAB74844	Aab74844 S. griseu
21	23	4.9	423	2	AAW68098	Aaw68098 Chlamydom
22	22	4.7	22	4	AAB67674	Aab67674 Amino aci
23	22	4.7	22	5	AAM49518	Aam49518 His-tag p
24	22	4.7	26	7	ADD31745	Add31745 Purificat
25	22	4.7	53	2	AAW36986	Aaw36986 HFKT pept
26	22	4.7	54	2	AAW36990	Aaw36990 HFKT pept
27	22	4.7	91	2	AAY09127	Aay09127 Recombina
28	22	4.7	93	2	AAY09128	Aay09128 Recombina
29	22	4.7	195	3	AAB11377	Aab11377 B. stearo
30	22	4.7	269	5	AAM49520	Aam49520 B. mori 1
31	22	4.7	269	5	AAM49541	Aam49541 B. mori T
32	22	4.7	323	2	AAW71185	Aaw71185 FP505 pro
33	22	4.7	332	4	AAE02191	Aae02191 Staphyloc
34	22	4.7	364	3	AAB13011	Aab13011 Q-tagged
35	22	4.7	567	4	AAU00220	Aau00220 Bad-DTTR
36	21	4.4	21	2	AAW23653	Aaw23653 Recombina
37	21	4.4	21	2	AAW23650	Aaw23650 Recombina
38	21	4.4	21	2	AAW23647	Aaw23647 Recombina
39	21	4.4	194	5	AAM51139	Aam51139 HIV TAT t
40	20	4.2	411	4	AAU00219	Aau00219 Bcl-Xl-DT
41	17	3.6	24	2	AAW19779	Aaw19779 Polyhisti
42	17	3.6	24	2	AAY06647	Aay06647 Polyhisti
43	17	3.6	24	2	AAY27015	Aay27015 Met Gly H
44	17	3.6	24	2	AAY09370	Aay09370 12 amino
45	17	3.6	24	3	AAY81959	Aay81959 Enterokin

ALIGNMENTS

RESULT 1
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant

```
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             100.0%;  Score 472;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 2
AAY77135
ID   AAY77135 standard; protein; 440 AA.
XX
AC   AAY77135;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype B (BoNTB) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype B; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87213.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 23; Page; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast. Note: This
CC   sequence is not given in the specification, but is decoded from the BoNTB
CC   Hc DNA sequence given on pages 39-40
XX
SQ   Sequence 440 AA;

  Query Match             14.8%;  Score 70;  DB 3;  Length 440;
  Best Local Similarity   100.0%;  Pred. No. 2e-57;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 150

Qy        183 WFFVTITNNL 192
              ||||||||||
Db        151 WFFVTITNNL 160


RESULT 3
AAB04091
ID   AAB04091 standard; protein; 440 AA.
XX
AC   AAB04091;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype B).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
```

PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54485.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 4b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 440 AA;

Query Match 14.8%; Score 70; DB 4; Length 440;
Best Local Similarity 100.0%; Pred. No. 2e-57;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 91 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 150

Qy 183 WFFVTITNNL 192
||||||||||
Db 151 WFFVTITNNL 160

RESULT 4
AAW68394
ID AAW68394 standard; protein; 472 AA.
XX
AC AAW68394;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30581.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 303-305; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 14.8%; Score 70; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182

Qy 183 WFFVTITNNL 192
||||||||||
Db 183 WFFVTITNNL 192

RESULT 5
AAY93308
ID AAY93308 standard; protein; 1070 AA.
XX
AC AAY93308;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype B.
XX
OS Synthetic.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.

XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 45-47; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype B. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a domain
CC that translocates the SOD of the composition into the neuronal cell.
CC After translocation, the linker is cleaved to release the SOD. The
CC composition is useful for treating neuronal diseases caused or augmented
CC by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC disease, Huntington's disease and motor neurone diseases
XX
SQ Sequence 1070 AA;

Query Match 14.8%; Score 70; DB 3; Length 1070;
Best Local Similarity 100.0%; Pred. No. 4.4e-57;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 780

Qy 183 WFFVTITNNL 192
||||||||||
Db 781 WFFVTITNNL 790

RESULT 6
AAY93311
ID AAY93311 standard; protein; 1095 AA.
XX
AC AAY93311;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype B.
XX
OS Synthetic.
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 54-57; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide dismutase
CC (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC serotype B. The specification describes a composition for delivery of SOD
CC to neuronal cells. The composition comprises SOD linked, by a cleavable
CC linker, to a neuronal cell targeting component (NCTC). This component has
CC a domain that binds to a neuronal cell and a domain that translocates the
CC SOD of the composition into the neuronal cell. After translocation, the
CC linker is cleaved to release the SOD. The composition is useful for
CC treating neuronal diseases caused or augmented by oxidative stress, such
CC as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases
XX
SQ Sequence 1095 AA;

Query Match 14.8%; Score 70; DB 3; Length 1095;
Best Local Similarity 100.0%; Pred. No. 4.5e-57;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 746 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 805

Qy 183 WFFVTITNNL 192
||||||||||
Db 806 WFFVTITNNL 815

RESULT 7
AAW56017
ID AAW56017 standard; protein; 1169 AA.
XX
AC AAW56017;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH728/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26289.
XX

PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS Example 2; Page 91-94; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 1169 AA;

Query Match 14.8%; Score 70; DB 2; Length 1169;
Best Local Similarity 100.0%; Pred. No. 4.8e-57;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 942 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy 183 WFFVTITNNL 192
||||||||||
Db 1002 WFFVTITNNL 1011

RESULT 8
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30579.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 291-296; 428pp; English.
XX
CC This is the amino acid sequence of the type B toxin of Clostridium
CC botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC B toxin has been expressed as a histidine-tagged protein in Escherichia
CC coli. The invention relates to C. botulinum recombinant toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of the recombinant proteins are E. coli, insect cells and
CC yeast cells. The recombinant proteins are used as immunogens for the
CC production of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1291 AA;

Query Match 14.8%; Score 70; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 5.2e-57;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 942 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy 183 WFFVTITNNL 192
||||||||||
Db 1002 WFFVTITNNL 1011

RESULT 9
AAB04082
ID AAB04082 standard; protein; 848 AA.
XX
AC AAB04082;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Page 13; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a

CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 848 AA;

Query Match 8.9%; Score 42; DB 4; Length 848;
Best Local Similarity 100.0%; Pred. No. 1.2e-30;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNL 192
||||||||||||||||||||||||||||||||||||||||||||
Db 528 RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNL 569

RESULT 10
AAB04085
ID AAB04085 standard; protein; 439 AA.
XX
AC AAB04085;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin C fragment sequence (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Example 8; Page 38; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 439 AA;

Query Match 8.3%; Score 39; DB 4; Length 439;
Best Local Similarity 100.0%; Pred. No. 4.5e-28;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNL 192
|||||||||||||||||||||||||||||||||||||
Db 121 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNL 159

RESULT 11
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30596.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC express BotG soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines

```
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match             6.6%;  Score 31;  DB 2;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31
              |||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31


RESULT 12
AAW68399
ID   AAW68399 standard; protein; 448 AA.
XX
AC   AAW68399;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type F toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW   botulism; BotF.
XX
OS   Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 48; Page 364-365; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC   sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC   express BotC soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 448 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 13
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
```

```
RESULT 14
AAW68398
ID   AAW68398 standard; protein; 451 AA.
XX
AC   AAW68398;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type D toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW   botulism; BotD.
XX
OS   Clostridium botulinum; serotype D.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30591.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 47; Page 352-353; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC   (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC   BotD soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 15
AAW68396
ID   AAW68396 standard; protein; 452 AA.
XX
AC   AAW68396;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain NCTC 11219.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


Search completed: September 10, 2004, 05:05:54
Job time : 30.6553 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 7.73869 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	70	14.8	1169	4	US-09-255-829-20	Sequence 20, Appl
2	26	5.5	462	1	US-08-480-604A-26	Sequence 26, Appl
3	26	5.5	462	2	US-08-405-496A-26	Sequence 26, Appl
4	26	5.5	462	3	US-08-915-136-26	Sequence 26, Appl
5	26	5.5	462	4	US-09-084-517-26	Sequence 26, Appl
6	23	4.9	23	1	US-08-480-604A-24	Sequence 24, Appl
7	23	4.9	23	2	US-08-405-496A-24	Sequence 24, Appl
8	23	4.9	23	3	US-08-915-136-24	Sequence 24, Appl
9	23	4.9	23	4	US-09-084-517-24	Sequence 24, Appl
10	22	4.7	22	4	US-09-660-742-3	Sequence 3, Appli
11	22	4.7	53	2	US-08-651-818A-19	Sequence 19, Appl
12	22	4.7	53	3	US-09-184-826-19	Sequence 19, Appl
13	22	4.7	54	2	US-08-651-818A-23	Sequence 23, Appl
14	22	4.7	54	3	US-09-184-826-23	Sequence 23, Appl
15	22	4.7	323	1	US-08-591-989-5	Sequence 5, Appli
16	21	4.4	21	2	US-08-651-818A-21	Sequence 21, Appl
17	21	4.4	21	3	US-09-184-826-21	Sequence 21, Appl
18	17	3.6	24	3	US-08-584-031-8	Sequence 8, Appli
19	17	3.6	24	3	US-08-780-496-8	Sequence 8, Appli
20	17	3.6	24	4	US-08-928-069-9	Sequence 9, Appli
21	15	3.2	20	4	US-09-186-489-6	Sequence 6, Appli
22	15	3.2	20	4	US-10-043-665B-6	Sequence 6, Appli
23	13	2.8	858	4	US-09-255-829-22	Sequence 22, Appl
24	13	2.8	858	4	US-09-255-829-29	Sequence 29, Appl
25	12	2.5	16	5	PCT-US94-04361-56	Sequence 56, Appl
26	12	2.5	17	2	US-08-600-783-13	Sequence 13, Appl
27	11	2.3	20	4	US-09-674-677-34	Sequence 34, Appl
28	11	2.3	24	4	US-09-292-858B-27	Sequence 27, Appl
29	11	2.3	254	3	US-09-167-434-7	Sequence 7, Appli
30	11	2.3	254	3	US-08-853-755-7	Sequence 7, Appli
31	11	2.3	302	2	US-08-203-532F-4	Sequence 4, Appli
32	11	2.3	302	3	US-08-950-860-16	Sequence 16, Appl
33	11	2.3	302	3	US-09-078-465-4	Sequence 4, Appli
34	11	2.3	302	5	PCT-US95-01882A-4	Sequence 4, Appli
35	11	2.3	303	2	US-08-203-532F-2	Sequence 2, Appli
36	11	2.3	303	3	US-09-078-465-2	Sequence 2, Appli
37	11	2.3	303	5	PCT-US95-01882A-2	Sequence 2, Appli
38	11	2.3	363	4	US-09-328-352-4930	Sequence 4930, Ap
39	11	2.3	414	5	PCT-US92-06840-2	Sequence 2, Appli
40	10	2.1	25	4	US-09-721-154-14	Sequence 14, Appl
41	10	2.1	335	3	US-09-095-117-2	Sequence 2, Appli
42	10	2.1	335	4	US-09-850-260-2	Sequence 2, Appli
43	10	2.1	449	2	US-08-927-394-2	Sequence 2, Appli
44	10	2.1	480	1	US-07-882-292-2	Sequence 2, Appli
45	10	2.1	480	2	US-08-331-644-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 14.8%; Score 70; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 1.3e-59;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        183 WFFVTITNNL 192
              ||||||||||
Db       1002 WFFVTITNNL 1011
```

RESULT 2
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-26

Query Match 5.5%; Score 26; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 4.3e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
```

RESULT 3
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-26

Query Match 5.5%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 4.3e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
```

RESULT 4
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:

```
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             5.5%;  Score 26;  DB 3;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 5
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             5.5%;  Score 26;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 6
US-08-480-604A-24
; Sequence 24, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
```

```
;       STREET:  220 MONTGOMERY STREET, SUITE 2200
;       CITY:  SAN FRANCISCO
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,604A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/422,711
;       FILING DATE:  14-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/405,496
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  INGOLIA, DIANE E.
;       REGISTRATION NUMBER:  40,027
;       REFERENCE/DOCKET NUMBER:  OPHD-01763
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  23 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-480-604A-24

  Query Match             4.9%;  Score 23;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 7
US-08-405-496A-24
; Sequence 24, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-24

  Query Match             4.9%;  Score 23;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 8
US-08-915-136-24
; Sequence 24, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
```

```
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-24

  Query Match             4.9%;  Score 23;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 9
US-09-084-517-24
; Sequence 24, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-24

  Query Match             4.9%;  Score 23;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 10
US-09-660-742-3
; Sequence 3, Application US/09660742
; Patent No. 6485977
; GENERAL INFORMATION:
;  APPLICANT:  Collmer, Alan
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  RECOMBINANT CONSTRUCTS AND TECHNIQUES FOR DELIVERING TO
;  TITLE OF INVENTION:  EUCARYOTIC CELLS BACTERIAL PROTEINS THAT ARE SECRETED
;  TITLE OF INVENTION:  VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE:  19603/4021
;  CURRENT APPLICATION NUMBER: US/09/660,742
;  CURRENT FILING DATE:  2000-09-13
;  PRIOR APPLICATION NUMBER:  60/153,507
;  PRIOR FILING DATE:  1999-09-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 22
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:* Polypeptide
US-09-660-742-3

  Query Match             4.7%;  Score 22;  DB 4;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 11
```

```
US-08-651-818A-19
; Sequence 19, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-19

  Query Match             4.7%;  Score 22;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 12
US-09-184-826-19
; Sequence 19, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-19

  Query Match             4.7%;  Score 22;  DB 3;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 13
US-08-651-818A-23
; Sequence 23, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-23

  Query Match             4.7%;  Score 22;  DB 2;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 14
US-09-184-826-23
; Sequence 23, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-23

  Query Match             4.7%;  Score 22;  DB 3;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 15
US-08-591-989-5
; Sequence 5, Application US/08591989
; Patent No. 5795721
;  GENERAL INFORMATION:
;    APPLICANT:  Ross S. Rabin, Sumedha Jayasena
;    APPLICANT:  and Larry Gold
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC
;    TITLE OF INVENTION:  ACID LIGANDS OF ICP4
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 East Prentice Avenue, Suite #200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.40 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX 49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-591-989-5

  Query Match             4.7%;  Score 22;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db         17 MGHHHHHHHHHHSSGHIEGRHM 38


Search completed: September 10, 2004, 05:17:04
Job time : 8.73869 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 23.876 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	472	100.0	472	12	US-10-272-898-44	Sequence 44, Appl
2	472	100.0	472	15	US-10-354-774-44	Sequence 44, Appl
3	472	100.0	472	15	US-10-271-012-44	Sequence 44, Appl
4	472	100.0	472	16	US-10-729-122-44	Sequence 44, Appl
5	472	100.0	472	16	US-10-729-039-44	Sequence 44, Appl
6	439	93.0	1291	12	US-10-272-898-40	Sequence 40, Appl
7	439	93.0	1291	15	US-10-354-774-40	Sequence 40, Appl
8	439	93.0	1291	15	US-10-271-012-40	Sequence 40, Appl
9	439	93.0	1291	15	US-10-452-024-117	Sequence 117, App
10	439	93.0	1291	15	US-10-205-516-4	Sequence 4, Appli
11	439	93.0	1291	16	US-10-729-122-40	Sequence 40, Appl
12	439	93.0	1291	16	US-10-729-039-40	Sequence 40, Appl
13	439	93.0	1301	15	US-10-205-516-18	Sequence 18, Appl
14	142	30.1	361	15	US-10-452-024-175	Sequence 175, App
15	142	30.1	361	15	US-10-452-024-176	Sequence 176, App
16	70	14.8	440	10	US-09-910-186A-8	Sequence 8, Appli
17	70	14.8	472	12	US-10-272-898-46	Sequence 46, Appl
18	70	14.8	472	15	US-10-354-774-46	Sequence 46, Appl
19	70	14.8	472	15	US-10-271-012-46	Sequence 46, Appl
20	70	14.8	472	16	US-10-729-122-46	Sequence 46, Appl
21	70	14.8	472	16	US-10-729-039-46	Sequence 46, Appl
22	70	14.8	1169	14	US-10-241-596-20	Sequence 20, Appl
23	70	14.8	1290	15	US-10-360-101-220	Sequence 220, App
24	70	14.8	1290	15	US-10-452-024-119	Sequence 119, App
25	70	14.8	1291	12	US-10-272-898-42	Sequence 42, Appl
26	70	14.8	1291	15	US-10-354-774-42	Sequence 42, Appl
27	70	14.8	1291	15	US-10-271-012-42	Sequence 42, Appl
28	70	14.8	1291	15	US-10-452-024-2	Sequence 2, Appli
29	70	14.8	1291	15	US-10-452-024-118	Sequence 118, App
30	70	14.8	1291	16	US-10-729-122-42	Sequence 42, Appl
31	70	14.8	1291	16	US-10-729-039-42	Sequence 42, Appl
32	45	9.5	441	15	US-10-452-024-188	Sequence 188, App
33	45	9.5	1291	15	US-10-452-024-142	Sequence 142, App
34	40	8.5	1291	15	US-10-452-024-121	Sequence 121, App
35	40	8.5	1291	15	US-10-452-024-122	Sequence 122, App
36	40	8.5	1291	15	US-10-452-024-123	Sequence 123, App
37	31	6.6	473	12	US-10-272-898-79	Sequence 79, Appl
38	31	6.6	473	15	US-10-354-774-79	Sequence 79, Appl
39	31	6.6	473	15	US-10-271-012-79	Sequence 79, Appl
40	31	6.6	473	16	US-10-729-122-79	Sequence 79, Appl
41	31	6.6	473	16	US-10-729-039-79	Sequence 79, Appl
42	26	5.5	448	12	US-10-272-898-73	Sequence 73, Appl
43	26	5.5	448	15	US-10-354-774-73	Sequence 73, Appl
44	26	5.5	448	15	US-10-271-012-73	Sequence 73, Appl
45	26	5.5	448	16	US-10-729-122-73	Sequence 73, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-44
; Sequence 44, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:

```
;          LENGTH: 472 amino acids
;          TYPE: amino acid
;          TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-272-898-44

  Query Match           100.0%;  Score 472;  DB 12;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 2
US-10-354-774-44
; Sequence 44, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-354-774-44

  Query Match           100.0%;  Score 472;  DB 15;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 3
US-10-271-012-44
; Sequence 44, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-271-012-44

  Query Match             100.0%;  Score 472;  DB 15;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 4
US-10-729-122-44
; Sequence 44, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A. .
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-729-122-44

  Query Match             100.0%;  Score 472;  DB 16;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 5
US-10-729-039-44
```

```
; Sequence 44, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 44:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-729-039-44

  Query Match           100.0%;  Score 472;  DB 16;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 6
US-10-272-898-40
; Sequence 40, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 40:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1291 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-272-898-40

  Query Match           93.0%;  Score 439;  DB 12;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
                 ||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291


RESULT 7
US-10-354-774-40
; Sequence 40, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-354-774-40

  Query Match             93.0%;  Score 439;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
                 ||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291


RESULT 8
US-10-271-012-40
; Sequence 40, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-271-012-40

  Query Match             93.0%;  Score 439;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291
```

RESULT 9
US-10-452-024-117
; Sequence 117, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 117
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-117

Query Match 93.0%; Score 439; DB 15; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291
```

RESULT 10
US-10-205-516-4
; Sequence 4, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-4

Query Match 93.0%; Score 439; DB 15; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291
```

```
Db       1273 KSNLGCNWQFIPKDEGWTE 1291


RESULT 11
US-10-729-122-40
; Sequence 40, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-729-122-40

  Query Match             93.0%;  Score 439;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              |||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291


RESULT 12
US-10-729-039-40
; Sequence 40, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-729-039-40

  Query Match             93.0%;  Score 439;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032
```

```
Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291
```

RESULT 13
US-10-205-516-18
; Sequence 18, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1301
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-18

Query Match 93.0%; Score 439; DB 15; Length 1301;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 916

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 976

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1036

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1037 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1096

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1097 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1156

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1157 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1216

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1217 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1276

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||
Db       1277 KSNLGCNWQFIPKDEGWTE 1295
```

RESULT 14
US-10-452-024-175
; Sequence 175, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 175
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-175

Query Match 30.1%; Score 142; DB 15; Length 361;
Best Local Similarity 100.0%; Pred. No. 5.1e-127;
Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 279

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 339

Qy        154 WTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||
Db        340 WTLIDINGKTKSVFFEYNIRED 361
```

RESULT 15
US-10-452-024-176
; Sequence 176, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 176
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-176

Query Match 30.1%; Score 142; DB 15; Length 361;
Best Local Similarity 100.0%; Pred. No. 5.1e-127;
Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 279
```

```
Qy          94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         280 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 339

Qy         154 WTLIDINGKTKSVFFEYNIRED 175
               |||||||||||||||||||||
Db         340 WTLIDINGKTKSVFFEYNIRED 361
```

Search completed: September 10, 2004, 05:52:33
Job time : 24.876 secs

This Page Blank (uspto)

```
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;186-242/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 4
A56837
homeotic protein MOX2 - human
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 24-Sep-1999
C;Accession: A56837; S52168
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.S.; Theodorakis, K.; Mankoo, B.; Pachnis, V.;
Genomics 26, 550-555, 1995
A;Title: Isolation of the human MOX2 homeobox gene and localization to chromosome 7p22.1
A;Reference number: A56837; MUID:95331791; PMID:7607679
A;Accession: A56837
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-303 <GRI>
A;Cross-references: EMBL:X82629; NID:g732790; PIDN:CAA57949.1; PID:g732791
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.; Theodorakis, K.; Maukoo, B.; Pachnis, V.; K
submitted to the EMBL Data Library, November 1994
A;Description: Isolation of the human MOX2 homeobox gene and localization to chromosome
A;Reference number: S52168
A;Accession: S52168
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-7,'R',9-110,'R',112-122,'Q',124-157,'V',159-303 <GR2>
A;Cross-references: EMBL:X82629
C;Genetics:
A;Gene: GDB:MOX2
A;Cross-references: GDB:120703; OMIM:155970
A;Map position: 3q13-3q13
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
B49122
homeobox protein Mox-2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Dec-1993 #sequence_revision 17-Nov-1995 #text_change 17-Nov-2000
C;Accession: B49122; S41779; S29902
R;Candia, A.F.; Hu, J.; Crosby, J.; Lalley, P.A.; Noden, D.; Nadeau, J.H.; Wright, C.V.
Development 116, 1123-1136, 1992
A;Title: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially e
A;Reference number: A49122; MUID:93201999; PMID:1363541
A;Accession: B49122
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN1>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
A;Experimental source: C57BL
A;Note: the complete translation is not shown
R;Candia, A.F.; Kovalik, J.P.; Wright, C.V.E.
Nucleic Acids Res. 21, 4982, 1993
A;Title: Amino acid sequence of Mox-2 and comparison to its Xenopus and rat homologs.
A;Reference number: S41469; MUID:94232829; PMID:7909944
A;Accession: S41779
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN2>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
R;Candia, A.
submitted to the EMBL Data Library, October 1992
A;Reference number: S29902
A;Accession: S29902
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 186-246 <CAN>
A;Cross-references: EMBL:Z16406
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 6
A48130
growth arrest-specific homeobox protein Cvx - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A48130; S31976
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
Mol. Cell. Biol. 13, 3722-3733, 1993
A;Title: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated du
A;Reference number: A48130; MUID:93268321; PMID:8098844
A;Accession: A48130
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-303 <GOR>
A;Cross-references: GB:Z17223; NID:g57951; PIDN:CAA78931.1; PID:g57952
A;Experimental source: aorta
A;Note: sequence extracted from NCBI backbone (NCBIN:132842, NCBIP:132843)
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
submitted to the EMBL Data Library, October 1992
A;Description: Cvx: A diverged homeodomain gene which is rapidly down-regulated followin
A;Reference number: S31976
A;Accession: S31976
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-182,'S',184-303 <GO2>
A;Cross-references: EMBL:Z17223
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 7
S39406
homeotic protein otx1 - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 15-Oct-1999
C;Accession: S39406
```

R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; B
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S39406
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-354 <SIM>
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 8
S35345
otx1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000
C;Accession: S35345; S29087
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; B
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S35345
A;Molecule type: DNA
A;Residues: 1-355 <SIM>
R;Simeone, A.; Acampora, D.; Gulisano, M.; Stornaiuolo, A.; Boncinelli, E.
Nature 358, 687-690, 1992
A;Title: Nested expression domains of four homeobox genes in developing rostral brain.
A;Reference number: S29084; MUID:92357147; PMID:1353865
A;Accession: S29087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 31-98 <SI2>
A;Cross-references: EMBL:X68883; NID:g53540; PIDN:CAA48754.1; PID:g53541
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 9
I56547
homeodomain protein otx1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C;Accession: I56547
R;Frantz, G.D.; Weimann, J.M.; Levin, M.E.; McConnell, S.K.
J. Neurosci. 14, 5725-5740, 1994
A;Title: Otx1 and Otx2 define layers and regions in developing cerebral cortex and cereb
A;Reference number: I56547; MUID:95016961; PMID:7931541
A;Accession: I56547
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: GB:L32602; NID:g535739; PIDN:AAA53557.1; PID:g535740
C;Genetics:
A;Gene: OTX1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 10
A40350
transcription repressor protein YY1 - human
N;Alternate names: transcription repressor/activator protein NF-E1
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Oct-1999
C;Accession: A40350; S78494; S33712; A56419
R;Shi, Y.; Seto, E.; Chang, L.S.; Shenk, T.
Cell 67, 377-388, 1991
A;Title: Transcriptional repression by YY1, a human GLI-Krueppel-related protein, and rel
A;Reference number: A40350; MUID:92005716; PMID:1655281
A;Accession: A40350
A;Molecule type: mRNA
A;Residues: 1-414 <SHI>
A;Cross-references: GB:M77698; NID:g186767; PIDN:AAA59467.1; PID:g186768
A;Experimental source: HeLa cells
A;Note: the authors translated the codon CGC for residue 371 as Lys, CGA for residue 375
R;Whitson, R.H.; Huang, T.; Dang, J.; Itakura, K.
submitted to the EMBL Data Library, July 1992
A;Description: Observed and predicted DNA binding of a zinc finger protein which recogni
A;Reference number: S78494
A;Accession: S78494
A;Molecule type: mRNA
A;Residues: 1-195,'G',197-414 <WHI>
A;Cross-references: EMBL:Z14077; NID:g38010; PIDN:CAA78455.1; PID:g38011
R;Park, K.; Atchison, M.L.
Proc. Natl. Acad. Sci. U.S.A. 88, 9804-9808, 1991
A;Title: Isolation of a candidate repressor/activator, NF-E1 (YY-1), delta, that binds to
A;Reference number: A56419; MUID:92052179; PMID:1946405
A;Accession: S33712
A;Molecule type: mRNA
A;Residues: 1-64,'R',66-195,'G',197-414 <PAR>
A;Cross-references: GB:M76541; NID:g189173; PIDN:AAA59926.1; PID:g189174
C;Keywords: DNA binding; transcription regulation; zinc finger
F;298-320/Region: zinc finger CCHH motif
F;327-347/Region: zinc finger CCHH motif
F;355-377/Region: zinc finger CCHH motif
F;385-407/Region: zinc finger CCHH motif

Query Match 2.3%; Score 11; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 69 GHHHHHHHHHH 79

RESULT 11
A48273
delta/YY1/NF-E1/UCRBP transcription factor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 01-Dec-2000
C;Accession: A48273; A42055; A56418
R;Safrany, G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 5559-5563, 1993

```
A;Title: Characterization of the mouse gene that encodes the delta/YY1/NF-E1/UCRBP trans
A;Reference number: A48273; MUID:93296177; PMID:8516301
A;Accession: A48273
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-414 <RES>
A;Cross-references: GB:L13968; NID:g293847; PIDN:AAA40477.1; PID:g293849
R;Flanagan, J.R.; Becker, K.G.; Ennist, D.L.; Gleason, S.L.; Driggers, P.H.; Levi, B.Z.;
Mol. Cell. Biol. 12, 38-44, 1992
A;Title: Cloning of a negative transcription factor that binds to the upstream conserved
A;Reference number: A42055; MUID:92107191; PMID:1309593
A;Accession: A42055
A;Molecule type: mRNA
A;Residues: 1-414 <FLA>
A;Cross-references: GB:M73963; NID:g202270; PIDN:AAA40522.1; PID:g202271
A;Note: sequence extracted from NCBI backbone (NCBIN:74641, NCBIP:74642)
R;Hariharan, N.; Kelley, D.E.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 88, 9799-9803, 1991
A;Title: delta, a transcription factor that binds to downstream elements in several poly
A;Reference number: A56418; MUID:92052178; PMID:1946404
A;Accession: A56418
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218,'S',220-374,'G',376-414 <HAR>
A;Cross-references: GB:M74590; NID:g192940; PIDN:AAA37521.1; PID:g192941
C;Genetics:
A;Introns: 227/1; 281/2; 301/3; 354/3
C;Keywords: transcription factor; zinc finger

  Query Match              2.3%;  Score 11;  DB 2;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0054;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 12
T39712
hypothetical protein SPBC17D11.01 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: hypothetical protein SPBC20F10.11
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Jan-2000
C;Accession: T39712; T39882
R;Beck, A.; Reinhardt, R.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21872
A;Accession: T39712
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-420 <BEC>
A;Cross-references: EMBL:AL031739; PIDN:CAA21072.1; GSPDB:GN00067; SPDB:SPBC17D11.01
A;Experimental source: strain 972h-; cosmid c17D11
R;Lyne, M.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21888
A;Accession: T39882
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-281 <LYN>
A;Cross-references: EMBL:AL021747; PIDN:CAA16851.1; GSPDB:GN00067
C;Genetics:
A;Gene: SPBC20F10.11; SPDB:SPBC17D11.01
A;Map position: 2
A;Introns: 10/1; 28/2
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match              2.3%;  Score 11;  DB 2;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.0054;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 13
T42516
hypothetical protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T42516
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42516
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-427 <YOS>
A;Cross-references: EMBL:D89156; NID:g1749519; PIDN:BAA13818.1; PID:g1749520
A;Experimental source: strain PR745
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match              2.3%;  Score 11;  DB 2;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 0.0055;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 14
S69205
stripe a/b protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S69205
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube gu
A;Reference number: S69205; MUID:96203082; PMID:8612588
A;Accession: S69205
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1180 <FRO>
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789
C;Keywords: alternative splicing

  Query Match              2.3%;  Score 11;  DB 2;  Length 1180;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db        641 HHHHHHHHHHS 651


RESULT 15
I39193
gene HOXA1 protein - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 15-Oct-1999
C;Accession: I39193
R;Chariot, A.; Moreau, L.; Senterre, G.; Sobel, M.; Castronovo, V.
Biochem. Biophys. Res. Commun. 215, 713-720, 1995
A;Title: Retinoic acid induces three newly cloned HOXA1 transcripts in MCF7 breast cance
A;Reference number: I39192; MUID:96011836; PMID:7488013
A;Accession: I39193
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-132 <RES>
A;Cross-references: EMBL:U37431; NID:g1051226; PIDN:AAC50250.1; PID:g1051229
C;Genetics:
```

A;Gene: GDB:HOXA1
A;Cross-references: GDB:120652; OMIM:142955
A;Map position: 7p15.3-7p15.3
A;Introns: 113/3
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 2.1%; Score 10; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 HHHHHHHHHH 12
              ||||||||||
Db         60 HHHHHHHHHH 69
```

Search completed: September 10, 2004, 05:14:45
Job time : 8.63889 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 6.53889 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	439	93.0	1291	2	I40631	non-proteolytic bo
2	70	14.8	1291	1	A48940	bontoxilysin (EC 3
3	11	2.3	302	2	A55641	homeotic protein G
4	11	2.3	303	2	A56837	homeotic protein M
5	11	2.3	303	2	B49122	homeobox protein M
6	11	2.3	303	2	A48130	growth arrest-spec
7	11	2.3	354	2	S39406	homeotic protein o
8	11	2.3	355	2	S35345	otx1 protein - mou
9	11	2.3	355	2	I56547	homeodomain protei
10	11	2.3	414	2	A40350	transcription repr
11	11	2.3	414	2	A48273	delta/YY1/NF-E1/UC
12	11	2.3	420	2	T39712	hypothetical prote
13	11	2.3	427	2	T42516	hypothetical prote
14	11	2.3	1180	2	S69205	stripe a/b protein
15	10	2.1	132	2	I39193	gene HOXA1 protein
16	10	2.1	133	2	B30242	stem cell protein
17	10	2.1	222	2	I39192	gene HOXA1 protein
18	10	2.1	236	2	S41512	Brn-3b protein - m
19	10	2.1	298	2	S41469	homeotic protein M
20	10	2.1	309	2	T29293	hypothetical prote
21	10	2.1	321	2	T02987	myb-related protei
22	10	2.1	326	2	D83483	probable metal tra
23	10	2.1	331	2	A30242	homeotic protein E
24	10	2.1	335	2	G01448	homeobox protein H
25	10	2.1	351	1	KGZQHL	histidine-rich gly
26	10	2.1	375	2	A46390	cAMP receptor subt
27	10	2.1	387	2	A47446	HNF-3/fork head fa
28	10	2.1	390	1	A38565	polycomb (Pc) prot
29	10	2.1	391	2	H86187	hypothetical prote
30	10	2.1	407	2	G84783	probable pectinest
31	10	2.1	410	2	I38502	gene Brn-3b protei
32	10	2.1	411	2	I58156	Brn-3.2 - mouse
33	10	2.1	430	2	S66671	neuron-derived rec
34	10	2.1	451	2	A55909	transforming prote
35	10	2.1	469	2	I37451	HBF-G2 (HFK-2) pro
36	10	2.1	476	2	A54743	transcription fact
37	10	2.1	480	2	JH0672	brain factor 1 pro
38	10	2.1	580	2	T46024	hypothetical prote
39	10	2.1	623	1	S33167	gene pointed prote
40	10	2.1	625	2	S71930	neuron-derived rec
41	10	2.1	628	2	JC2493	neuron derived orp
42	10	2.1	633	1	A26030	serine/threonine-s
43	10	2.1	657	2	A29454	knob-associated hi
44	10	2.1	665	2	S02358	lamin A - African
45	10	2.1	717	2	S38177	SSV7 protein homol

ALIGNMENTS

RESULT 1
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 93.0%; Score 439; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy          94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              |||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291
```

RESULT 2
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisio
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match   14.8%; Score 70; DB 1; Length 1291;
  Best Local Similarity 100.0%; Pred. No. 3.5e-64;
  Matches   70; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        183 WFFVTITNNL 192
              ||||||||||
Db       1002 WFFVTITNNL 1011
```

RESULT 3
A55641
homeotic protein GAX - human
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 24-Sep-1999
C;Accession: A55641
R;LePage, D.F.; Altomare, D.A.; Testa, J.R.; Walsh, K.
Genomics 24, 535-540, 1994
A;Title: Molecular cloning and localization of the human GAX gene to 7p21.
A;Reference number: A55641; MUID:95229154; PMID:7713505
A;Accession: A55641
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <LEP>
A;Cross-references: GB:L36328; NID:g557549; PIDN:AAA58497.1; PID:g557550
C;Genetics:
A;Gene: GAX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 4.07931 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	70	14.8	1290	1	BXB_CLOBO	P10844 clostridium
2	11	2.3	303	1	MOX2_HUMAN	P50222 homo sapien
3	11	2.3	303	1	MOX2_MOUSE	P32443 mus musculu
4	11	2.3	303	1	MOX2_RAT	P39020 rattus norv
5	11	2.3	354	1	OTX1_HUMAN	P32242 homo sapien
6	11	2.3	355	1	OTX1_MOUSE	P80205 mus musculu
7	11	2.3	355	1	OTX1_RAT	Q63410 rattus norv
8	11	2.3	414	1	TYY1_HUMAN	P25490 homo sapien
9	11	2.3	414	1	TYY1_MOUSE	Q00899 mus musculu
10	11	2.3	420	1	YBE1_SCHPO	O42980 schizosacch
11	11	2.3	466	1	ZIC3_MOUSE	Q62521 mus musculu
12	11	2.3	467	1	ZIC3_HUMAN	O60481 homo sapien
13	11	2.3	611	1	SNF1_CANGA	Q00372 candida gla
14	11	2.3	619	1	SNF1_CANTR	O94168 candida tro
15	10	2.1	202	1	HAN1_CHICK	Q90691 gallus gall
16	10	2.1	204	1	HAN1_SHEEP	Q28555 ovis aries
17	10	2.1	298	1	MOX2_XENLA	P39021 xenopus lae
18	10	2.1	331	1	HXA1_MOUSE	P09022 mus musculu
19	10	2.1	335	1	HXA1_HUMAN	P49639 homo sapien
20	10	2.1	351	1	HRPX_PLALO	P04929 plasmodium
21	10	2.1	375	1	CAR2_DICDI	P34907 dictyosteli
22	10	2.1	387	1	QIN_AVIS3	P56260 avian sarco
23	10	2.1	390	1	PC_DROME	P26017 drosophila
24	10	2.1	410	1	PO42_HUMAN	Q12837 homo sapien
25	10	2.1	411	1	PO42_MOUSE	Q63934 mus musculu
26	10	2.1	430	1	NOR2_RAT	Q63516 rattus norv
27	10	2.1	449	1	GAT6_HUMAN	Q92908 homo sapien
28	10	2.1	451	1	FXGB_CHICK	Q90964 gallus gall
29	10	2.1	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
30	10	2.1	465	1	HNF6_MOUSE	O08755 mus musculu
31	10	2.1	465	1	HNF6_RAT	P70512 rattus norv
32	10	2.1	469	1	FXGA_HUMAN	P55316 homo sapien
33	10	2.1	477	1	FXGB_HUMAN	P55315 homo sapien
34	10	2.1	480	1	FXGB_RAT	Q00939 rattus norv
35	10	2.1	481	1	FXGB_MOUSE	Q60987 mus musculu
36	10	2.1	551	1	CBX4_MOUSE	O55187 mus musculu
37	10	2.1	558	1	CBX4_HUMAN	O00257 homo sapien
38	10	2.1	620	1	SNF1_CANAL	P52497 candida alb
39	10	2.1	623	1	PNT1_DROME	P51022 drosophila
40	10	2.1	626	1	NR43_HUMAN	Q92570 homo sapien
41	10	2.1	628	1	NR43_RAT	P51179 rattus norv
42	10	2.1	633	1	SNF1_YEAST	P06782 saccharomyc
43	10	2.1	657	1	KNOB_PLAFN	P06719 plasmodium
44	10	2.1	665	1	LAMA_XENLA	P11048 xenopus lae
45	10	2.1	717	1	UBPB_YEAST	P36026 saccharomyc

ALIGNMENTS

RESULT 1
BXB_CLOBO
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence.";
RL Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.";
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;

```
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match           14.8%;  Score 70;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-63;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1000

Qy        183 WFFVTITNNL 192
              ||||||||||
Db       1001 WFFVTITNNL 1010


RESULT 2
MOX2_HUMAN
ID   MOX2_HUMAN     STANDARD;      PRT;   303 AA.
AC   P50222; Q9UPL6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE   homeobox).
GN   MEOX2 OR MOX2 OR GAX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95331791; PubMed=7607679;
RA   Grigoriou M., Kastrinaki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA   Pachnis V., Karagogeos D.;
RT   "Isolation of the human MOX2 homeobox gene and localization to
RT   chromosome 7p22.1-p21.3.";
RL   Genomics 26:550-555(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95229154; PubMed=7713505;
RA   Lepage D.F., Walsh K.;
RT   "Molecular cloning and localization of the human GAX gene to 7p21.";
RL   Genomics 24:535-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 230-303 FROM N.A.
RA   Cordes M., Lacy M.;
```

```
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Embryo and placenta.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82629; CAA57949.1; -.
DR   EMBL; L36328; AAA58497.1; -.
DR   EMBL; BC017021; AAH17021.1; -.
DR   EMBL; AC004452; AAC06184.1; -.
DR   PIR; A55641; A55641.
DR   PIR; A56837; A56837.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04005; -.
DR   Genew; HGNC:7014; MEOX2.
DR   MIM; 600535; -.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DNA_BIND    186    245       HOMEOBOX.
FT   CONFLICT     58     58       G -> D (IN REF. 2).
FT   CONFLICT     79     79       MISSING (IN REF. 2).
SQ   SEQUENCE   303 AA;  33457 MW;  809ADE0CD090023D CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 3
MOX2_MOUSE
ID   MOX2_MOUSE     STANDARD;      PRT;   303 AA.
AC   P32443;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201999; PubMed=1363541;
RA   Candia A.F., Hu J., Crosby J., Lalley P.A., Noden D., Nadeau J.H.,
RA   Wright C.V.E.;
RT   "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are
RT   differentially expressed during early mesodermal patterning in mouse
RT   embryos.";
RL   Development 116:1123-1136(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94232829; PubMed=7909944;
RA   Candia A.F., Kovalik J.-P., Wright C.V.E.;
RT   "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat
RT   homologs.";
RL   Nucleic Acids Res. 21:4982-4982(1993).
RN   [3]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=95349593; PubMed=7623821;
RA   Andres V., Fisher S., Wearsch P., Walsh K.;
RT   "Regulation of Gax homeobox gene transcription by a combination of
RT   positive factors including myocyte-specific enhancer factor 2.";
RL   Mol. Cell. Biol. 15:4272-4281(1995).
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-
CC       8.5 dpc it is found on the entire epithelium of the somite. At 9.5
CC       dpc its expression is restricted to the sclerotome. At 10.5 dpc it
CC       is found in sclerotomally derived cells including the vertebral
CC       and costal precursors.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16406; CAA78899.1; -.
DR   EMBL; S79168; -; NOT_ANNOTATED_CDS.
DR   PIR; B49122; B49122.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04048; -.
DR   MGD; MGI:103219; Meox2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       63     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33506 MW;  41BD05FC39AA4427 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 4
```

```
MOX2_RAT
ID   MOX2_RAT       STANDARD;      PRT;   303 AA.
AC   P39020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Growth arrest-specific homeobox).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93268321; PubMed=8098844;
RA   Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,
RA   Walsh K.;
RT   "Molecular cloning of a diverged homeobox gene that is rapidly down-
RT   regulated during the G0/G1 transition in vascular smooth muscle
RT   cells.";
RL   Mol. Cell. Biol. 13:3722-3733(1993).
RN   [2]
RP   REVISIONS.
RA   Walsh K.;
RL   Submitted (MAR-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Aorta and heart. Also detected in lung and
CC       kidney.
CC   -!- INDUCTION: Rapidly and transiently down-regulated during the
CC       transition from G0 to G1 induced by mitogen stimulation.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z17223; CAA78931.1; -.
DR   PIR; A48130; A48130.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04054; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       64     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33605 MW;  7776642AEFA3A2E8 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77
```

```
RESULT 5
OTX1_HUMAN
ID   OTX1_HUMAN     STANDARD;      PRT;   354 AA.
AC   P32242;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS):
CC       5'-TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC007621; AAH07621.1; -.
DR   PIR; S39406; S39406.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02079; -.
DR   Genew; HGNC:8521; OTX1.
DR   MIM; 600036; -.
DR   GO; GO:0005634; C:nucleus; NAS.
```

```
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    301       HIS-RICH.
SQ   SEQUENCE   354 AA;  37327 MW;  E32C1E47464BDDAA CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.0031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 6
OTX1_MOUSE
ID   OTX1_MOUSE     STANDARD;      PRT;   355 AA.
AC   P80205;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1 OR OTX-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 31-98 FROM N.A.
RX   MEDLINE=92357147; PubMed=1353865;
RA   Simeone A., Acampora D., Gulisano M., Stornaiuolo A., Boncinelli E.;
RT   "Nested expression domains of four homeobox genes in developing
RT   rostral brain.";
RL   Nature 358:687-690(1992).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC057105; AAH57105.1; -.
DR   EMBL; BC058354; AAH58354.1; -.
DR   EMBL; X68883; CAA48754.1; -.
DR   PIR; S35345; S35345.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02080; -.
DR   MGD; MGI:97450; Otx1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
SQ   SEQUENCE   355 AA;  37531 MW;  DA5136D9D4FC948F CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 7
OTX1_RAT
ID   OTX1_RAT       STANDARD;      PRT;   355 AA.
AC   Q63410; Q64203;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95016961; PubMed=7931541;
RA   Frantz G.D., Weimann J.M., Levin M.E., McConnell S.K.;
RT   "Otx1 and Otx2 define layers and regions in developing cerebral
RT   cortex and cerebellum.";
RL   J. Neurosci. 14:5725-5740(1994).
RN   [2]
RP   SEQUENCE OF 176-323 FROM N.A.
RX   MEDLINE=96108898; PubMed=8613727;
RA   Robel L., Ding M., James A.J., Lin X., Simeone A., Leckman J.F.,
RA   Vaccarino F.M.;
RT   "Fibroblast growth factor 2 increases Otx2 expression in precursor
RT   cells from mammalian telencephalon.";
RL   J. Neurosci. 15:7879-7891(1995).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3' (By similarity). May play a role in the specification
CC       or differentiation of neurons in the deep layers of the cerebral
CC       cortex, and also in cerebellar regionalization during early
CC       development.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: Expressed throughout the forebrain and
CC       midbrain during development, and in addition is also seen in
CC       discrete spatial and temporal domains in the developing cerebral
CC       cortex and cerebellum. Confined to a subpopulation of neurons in
CC       layers 5 and 6 within the adult cerebral cortex and during
CC       development expression is high in the progenitors of these deep-
CC       layer cells. Expressed in the developing cerebellum in spatially
CC       restricted regions of the external granular layer.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32602; AAA53557.1; -.
DR   EMBL; S81924; -; NOT_ANNOTATED_CDS.
DR   PIR; I56547; I56547.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
FT   CONFLICT    219    219       T -> A (IN REF. 2).
FT   CONFLICT    229    229       R -> G (IN REF. 2).
SQ   SEQUENCE   355 AA;  37602 MW;  C875871723D0B876 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
TYY1_HUMAN
ID   TYY1_HUMAN     STANDARD;      PRT;   414 AA.
AC   P25490; Q14935;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1).
GN   YY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92005716; PubMed=1655281;
RA   Shi Y., Seto E., Chang L.-S., Shenk T.;
RT   "Transcriptional repression by YY1, a human GLI-Kruppel-related
RT   protein, and relief of repression by adenovirus E1A protein.";
RL   Cell 67:377-388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=92052179; PubMed=1946405;
RA   Park K., Atchison M.;
RT   "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
RT   that binds to the immunoglobulin kappa 3' enhancer and the
RT   immunoglobulin heavy-chain mu E1 site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Whitson R.H., Huang T., Dang J., Itakura K.;
RL   Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98152931; PubMed=9493912;
RA   McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
RA   Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
RT   "Targeting of the YY1 transcription factor to the nucleolus and the
RT   nuclear matrix in situ: the C-terminus is a principal determinant for
RT   nuclear trafficking.";
RL   J. Cell. Biochem. 68:500-510(1998).
RN   [6]
```

```
RP   INTERACTION WITH YAF2.
RX   MEDLINE=97169297; PubMed=9016636;
RA   Kalenik J.L., Chen D., Bradley M.E., Chen S.-J., Lee T.-C.;
RT   "Yeast two-hybrid cloning of a novel zinc finger protein that
RT   interacts with the multifunctional transcription factor YY1.";
RL   Nucleic Acids Res. 25:843-849(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
RX   MEDLINE=97098436; PubMed=8942976;
RA   Houbaviy H.B., Usheva A., Shenk T., Burley S.K.;
RT   "Cocrystal structure of YY1 bound to the adeno-associated virus P5
RT   initiator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 353-379.
RX   MEDLINE=98308000; PubMed=9642075;
RA   Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
RA   Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
RT   "Design, synthesis and structure of a zinc finger with an artificial
RT   beta-turn.";
RL   J. Mol. Biol. 279:973-986(1998).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
CC       EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
CC       PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77698; AAA59467.1; -.
DR   EMBL; M76541; AAA59926.1; -.
DR   EMBL; Z14077; CAA78455.1; -.
DR   EMBL; BC037308; AAH37308.1; -.
DR   PIR; A40350; A40350.
DR   PDB; 1UBD; 23-DEC-96.
DR   PDB; 1ZNM; 01-APR-98.
DR   TRANSFAC; T00915; -.
DR   Genew; HGNC:12856; YY1.
DR   MIM; 600013; -.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     69       GLY-RICH.
FT   DOMAIN       70     80       POLY-HIS.
FT   DOMAIN      159    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE 1.
FT   ZN_FING     325    347       C2H2-TYPE 2.
FT   ZN_FING     353    377       C2H2-TYPE 3.
FT   ZN_FING     383    407       C2H2-TYPE 4.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION.
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN.
FT   CONFLICT     65     65       H -> R (IN REF. 2).
FT   CONFLICT    196    196       G -> R (IN REF. 1).
FT   STRAND      296    297
FT   TURN        301    302
FT   STRAND      306    307
FT   HELIX       310    317
FT   HELIX       318    320
FT   STRAND      325    326
FT   TURN        328    330
FT   STRAND      333    334
FT   HELIX       337    343
FT   HELIX       344    346
FT   TURN        347    347
FT   STRAND      353    354
FT   TURN        358    359
FT   STRAND      363    364
FT   HELIX       367    378
FT   TURN        388    389
FT   TURN        397    400
FT   HELIX       401    407
SQ   SEQUENCE   414 AA;  44712 MW;  058C05A0AD2D04E6 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0036;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 9
TYY1_MOUSE
ID   TYY1_MOUSE     STANDARD;      PRT;   414 AA.
AC   Q00899;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1) (UCR-motif DNA-binding protein).
GN   YY1 OR UCRBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107191; PubMed=1309593;
RA   Flanagan J.R., Becker K.G., Ennist D.L., Gleason S.L., Driggers P.H.,
RA   Levi B.-Z., Appella E., Ozato K.;
RT   "Cloning of a negative transcription factor that binds to the
RT   upstream conserved region of Moloney murine leukemia virus.";
RL   Mol. Cell. Biol. 12:38-44(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93296177; PubMed=8516301;
RA   Safrany G., Perry R.P.;
RT   "Characterization of the mouse gene that encodes the delta/YY1/NF-
RT   E1/UCRBP transcription factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5559-5563(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052178; PubMed=1946404;
RA   Hariharan N., Kelley D.E., Perry R.P.;
```

```
RT   "Delta, a transcription factor that binds to downstream elements in
RT   several polymerase II promoters, is a functionally versatile zinc
RT   finger protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9799-9803(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. BINDS TO
CC       THE UPSTREAM CONSERVED REGION (UCR) (5'-CGCCATTTT-3') OF MOLONEY
CC       MURINE LEUKEMIA VIRUS (MULV).
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73963; AAA40522.1; -.
DR   EMBL; L13968; AAA40477.1; -.
DR   EMBL; L13969; AAA40477.1; JOINED.
DR   EMBL; L13965; AAA40477.1; JOINED.
DR   EMBL; L13966; AAA40477.1; JOINED.
DR   EMBL; L13967; AAA40477.1; JOINED.
DR   EMBL; M74590; AAA37521.1; -.
DR   EMBL; BC055899; AAH55899.1; -.
DR   PIR; A48273; A48273.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T00278; -.
DR   TRANSFAC; T00865; -.
DR   MGD; MGI:99150; Yy1.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003700; F:transcription factor activity; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     70       GLY-RICH.
FT   DOMAIN       71     82       POLY-HIS.
FT   DOMAIN      161    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    407       C2H2-TYPE.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION
FT                                (BY SIMILARITY).
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION (BY SIMILARITY).
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN (BY SIMILARITY).
FT   CONFLICT    219    219       F -> S (IN REF. 3).
FT   CONFLICT    375    375       R -> G (IN REF. 3).
SQ   SEQUENCE   414 AA;  44717 MW;  C012378288E984F9 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0036;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 10
YBE1_SCHPO
ID   YBE1_SCHPO     STANDARD;      PRT;   420 AA.
AC   O42980; P78807;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative thiol protease C17D11.01 (EC 3.4.22.-).
GN   SPBC17D11.01 OR SPBC20F10.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- SIMILARITY: Belongs to peptidase family C48.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031739; CAA21072.1; -.
DR   EMBL; AL021747; CAA16851.1; -.
DR   EMBL; D89156; BAA13818.1; -.
DR   GeneDB_SPombe; SPBC17D11.01; -.
DR   InterPro; IPR003653; Peptidase_C48.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Thiol protease.
FT   DOMAIN      299    312       POLY-HIS.
FT   CONFLICT     57     57       L -> W (IN REF. 2).
FT   CONFLICT    420    420       F -> FGCYTCVV (IN REF. 2).
SQ   SEQUENCE   420 AA;  47257 MW;  2125D94D9A2D9E87 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.0037;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 11
ZIC3_MOUSE
ID   ZIC3_MOUSE     STANDARD;      PRT;   466 AA.
AC   Q62521;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=96132843; PubMed=8557628;
RA   Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA   Chapman V.M., Mikoshiba K.;
RT   "The mouse zic gene family. Homologues of the Drosophila pair-rule
RT   gene odd-paired.";
RL   J. Biol. Chem. 271:1043-1047(1996).
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: CNS. A high level expression is seen in the
CC       cerebellum.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D70849; BAA11116.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T04671; -.
DR   MGD; MGI:106676; Zic3.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     96       POLY-HIS.
FT   ZN_FING     294    321       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     327    351       C2H2-TYPE 2.
FT   ZN_FING     357    381       C2H2-TYPE 3.
FT   ZN_FING     387    409       C2H2-TYPE 4.
SQ   SEQUENCE   466 AA;  50644 MW;  089749DF726F069B CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 12
ZIC3_HUMAN
ID   ZIC3_HUMAN     STANDARD;      PRT;   467 AA.
AC   O60481;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HTX1 MET-323.
RX   MEDLINE=98016414; PubMed=9354794;
RA   Gebbia M., Ferrero G.B., Pilia G., Bassi M.T., Aylsworth A.S.,
RA   Penman-Splitt M., Bird L.M., Bamforth J.S., Burn J., Schlessiner D.,
RA   Nelson D.L., Casey B.;
RT   "X-linked situs abnormalities result from mutations in ZIC3.";
RL   Nat. Genet. 17:305-308(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in ZIC3 are the cause of X-Linked visceral
CC       heterotaxy (HTX1) [MIM:306955]. HTX1 is a disease characterized by
CC       congenital heart disease and alterations of visceral situs,
CC       including asplenia or polysplenia, symmetric liver, intestinal
CC       malrotation, and abnormal lung lobation.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF028706; AAC05594.1; -.
DR   EMBL; AL035443; CAB41648.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:12874; ZIC3.
DR   MIM; 300265; -.
DR   MIM; 306955; -.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein; Disease mutation.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     97       POLY-HIS.
FT   ZN_FING     295    322       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     328    352       C2H2-TYPE 2.
FT   ZN_FING     358    382       C2H2-TYPE 3.
FT   ZN_FING     388    410       C2H2-TYPE 4.
FT   VARIANT     323    323       T -> M (in HTX1).
FT                                /FTId=VAR_007753.
SQ   SEQUENCE   467 AA;  50569 MW;  3150CF13C0679568 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 0.004;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 13
SNF1_CANGA
ID   SNF1_CANGA     STANDARD;      PRT;   611 AA.
AC   Q00372;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCCLS84;
RX   MEDLINE=97101049; PubMed=8945576;
RA   Petter R., Kwon-Chung K.J.;
RT   "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT   pathogenic yeast Candida glabrata.";
RL   Infect. Immun. 64:5269-5273(1996).
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L78130; AAB48642.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       6     17       POLY-HIS.
FT   DOMAIN      39    290       PROTEIN KINASE.
FT   NP_BIND     45     53       ATP (BY SIMILARITY).
FT   BINDING     68     68       ATP (BY SIMILARITY).
FT   ACT_SITE   161    161       BY SIMILARITY.
FT   MOD_RES    194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  70049 MW;  89E17812A4900CD0 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 0.0051;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db          8 HHHHHHHHHHS 18


RESULT 14
SNF1_CANTR
ID   SNF1_CANTR     STANDARD;      PRT;   619 AA.
AC   O94168;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanai T., Ogawa K., Ueda M., Tanaka A.;
RT   "Genetic evaluation of the function of SNF1 in Candida tropicalis.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AB024535; BAA75889.1; -.
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000719; Prot_kinase.
```

```
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       17     29       POLY-HIS.
FT   DOMAIN       52    303       PROTEIN KINASE.
FT   NP_BIND      58     66       ATP (BY SIMILARITY).
FT   BINDING      81     81       ATP (BY SIMILARITY).
FT   ACT_SITE    174    174       BY SIMILARITY.
FT   MOD_RES     207    207       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  70323 MW;  0FCF1FC3DCE706D7 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 0.0052;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db         20 HHHHHHHHHHS 30


RESULT 15
HAN1_CHICK
ID   HAN1_CHICK     STANDARD;      PRT;   202 AA.
AC   Q90691;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Heart- and neural crest derivatives-expressed protein 1
DE   (Extraembryonic tissues, heart, autonomic nervous sytem and neural
DE   crest derivatives-expressed protein 1) (eHAND).
GN   HAND1 OR EHAND.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=96123273; PubMed=8533092;
RA   Srivastava D., Cserjesi P., Olson E.N.;
RT   "A subclass of bHLH proteins required for cardiac morphogenesis.";
RL   Science 270:1995-1998(1995).
CC   -!- FUNCTION: Plays an essential role in cardiac morphogenesis.
CC   -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC       bHLH protein.
CC   -!- DEVELOPMENTAL STAGE: At stage 8, detected in the cardiac crescent.
CC       At stage 10, expressed thoughout the cardiac tube and the sinus
CC       venosus. By stage 15, expressed homogenously in the various
CC       regions of the heart, including the atria, future left ventricle,
CC       bulbus cordis and truncus arteriosus, and in the forming branchial
CC       arches. Expression persists through stage 20, but decreases
CC       thereafter.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U40041; AAC59734.1; -.
DR   TRANSFAC; T04371; -.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS50888; HLH; 1.
KW   Developmental protein; Transcription regulation; DNA-binding;
KW   Nuclear protein.
FT   DOMAIN       9     18       POLY-HIS.
FT   DNA_BIND     82    95       BASIC DOMAIN.
FT   DOMAIN       96    139      HELIX-LOOP-HELIX MOTIF.
SQ   SEQUENCE   202 AA;  22679 MW;  1819896705268AA6 CRC64;

  Query Match             2.1%;  Score 10;  DB 1;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHH 12
              ||||||||||
Db          9 HHHHHHHHHH 18


Search completed: September 10, 2004, 05:07:12
Job time : 6.07931 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 19.1967 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	439	93.0	1291	2	Q08077	Q08077 clostridium
2	142	30.1	361	2	Q45848	Q45848 clostridium
3	142	30.1	361	2	Q45846	Q45846 clostridium
4	45	9.5	441	2	Q9X708	Q9x708 clostridium
5	45	9.5	1291	2	Q8GR96	Q8gr96 clostridium
6	40	8.5	1291	2	Q9ZAJ8	Q9zaj8 clostridium
7	40	8.5	1291	2	Q93G71	Q93g71 clostridium
8	40	8.5	1291	2	Q933K0	Q933k0 clostridium
9	26	5.5	78	2	Q7WUH9	Q7wuh9 clostridium
10	26	5.5	78	2	Q7WRW0	Q7wrw0 clostridium
11	12	2.5	226	11	Q9QZD2	Q9qzd2 rattus norv
12	12	2.5	516	13	O42186	O42186 brachydanio
13	12	2.5	517	13	Q90YL6	Q90yl6 brachydanio
14	12	2.5	569	13	Q802D2	Q802d2 brachydanio
15	12	2.5	679	13	Q803K9	Q803k9 brachydanio
16	12	2.5	706	5	Q8SXG1	Q8sxg1 drosophila
17	12	2.5	861	5	Q9VRT8	Q9vrt8 drosophila
18	11	2.3	163	5	Q7YTA3	Q7yta3 glomeris ma
19	11	2.3	225	13	Q90XZ3	Q90xz3 erpetoichth
20	11	2.3	230	4	O75263	O75263 homo sapien
21	11	2.3	243	4	Q9P2R2	Q9p2r2 homo sapien
22	11	2.3	243	6	Q9N2G2	Q9n2g2 gorilla gor
23	11	2.3	243	6	Q9N2G3	Q9n2g3 pan troglod
24	11	2.3	243	6	Q9N2G1	Q9n2g1 pongo pygma
25	11	2.3	302	13	Q90YH7	Q90yh7 gallus gall
26	11	2.3	303	6	Q95JA6	Q95ja6 sus scrofa
27	11	2.3	303	11	Q99M23	Q99m23 mus musculu
28	11	2.3	355	11	Q91ZJ3	Q91zj3 mus musculu
29	11	2.3	358	13	Q9DG32	Q9dg32 xenopus lae
30	11	2.3	411	11	Q8CHJ4	Q8chj4 rattus norv
31	11	2.3	441	5	Q8T1W3	Q8t1w3 dictyosteli
32	11	2.3	449	10	Q8L4M9	Q8l4m9 oryza sativ
33	11	2.3	456	11	Q8BSB3	Q8bsb3 mus musculu
34	11	2.3	467	13	Q9YGC6	Q9ygc6 xenopus lae
35	11	2.3	511	10	Q8GXZ5	Q8gxz5 arabidopsis
36	11	2.3	906	5	Q8IN91	Q8in91 drosophila
37	11	2.3	1122	5	Q8SXD4	Q8sxd4 drosophila
38	11	2.3	1147	5	Q95PH8	Q95ph8 dictyosteli
39	11	2.3	1180	5	Q24163	Q24163 drosophila
40	11	2.3	1186	5	Q9VED3	Q9ved3 drosophila
41	11	2.3	1283	5	Q95RH4	Q95rh4 drosophila
42	11	2.3	1911	5	Q9W4M7	Q9w4m7 drosophila
43	10	2.1	53	2	O55244	O55244 escherichia
44	10	2.1	233	4	Q8TC53	Q8tc53 homo sapien
45	10	2.1	233	5	Q8MP30	Q8mp30 dictyosteli

ALIGNMENTS

RESULT 1
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR PIR; I40631; I40631.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match             93.0%;  Score 439;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 912

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 972

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 1032

Qy        214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 1092

Qy        274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 1152

Qy        334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 1212

Qy        394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 1272

Qy        454 KSNLGCNWQFIPKDEGWTE 472
              |||||||||||||||||||
Db       1273 KSNLGCNWQFIPKDEGWTE 1291


RESULT 2
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;

  Query Match             30.1%;  Score 142;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-139;
  Matches  142;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 279

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 339

Qy        154 WTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||
Db        340 WTLIDINGKTKSVFFEYNIRED 361


RESULT 3
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;

  Query Match             30.1%;  Score 142;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-139;
  Matches  142;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 279

Qy         94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 339

Qy        154 WTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||
Db        340 WTLIDINGKTKSVFFEYNIRED 361


RESULT 4
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
```

```
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;

  Query Match             9.5%;  Score 45;  DB 2;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-38;
  Matches   45;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        249 QSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 293
              |||||||||||||||||||||||||||||||||||||||||||||
Db        226 QSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 270


RESULT 5
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             9.5%;  Score 45;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-37;
  Matches   45;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        249 QSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 293
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1068 QSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 1112


RESULT 6
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match             8.5%;  Score 40;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-32;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        254 EIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 293
              ||||||||||||||||||||||||||||||||||||||||
Db       1073 EIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 1112


RESULT 7
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
```

```
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             8.5%;  Score 40;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-32;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        254 EIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 293
              |||||||||||||||||||||||||||||||||||||||||
Db       1073 EIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 1112


RESULT 8
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             8.5%;  Score 40;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-32;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        254 EIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 293
              |||||||||||||||||||||||||||||||||||||||||
Db       1073 EIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 1112


RESULT 9
Q7WUH9
ID   Q7WUH9      PRELIMINARY;      PRT;    78 AA.
AC   Q7WUH9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CL196;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327855; AAQ16536.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE   78 AA;  9386 MW;  6E6465A3A32CBDBC CRC64;

  Query Match             5.5%;  Score 26;  DB 2;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1e-18;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        305 RSKYNQNSNYINYRNLYIGEKFIIRR 330
              ||||||||||||||||||||||||||
Db         34 RSKYNQNSNYINYRNLYIGEKFIIRR 59


RESULT 10
Q7WRW0
ID   Q7WRW0      PRELIMINARY;      PRT;    78 AA.
AC   Q7WRW0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=215, and 216;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327856; AAQ16537.1; -.
DR   EMBL; AY327857; AAQ16538.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE   78 AA;  9370 MW;  7CF465A3A32CBDBC CRC64;

  Query Match             5.5%;  Score 26;  DB 2;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1e-18;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        305 RSKYNQNSNYINYRNLYIGEKFIIRR 330
              ||||||||||||||||||||||||||
Db         34 RSKYNQNSNYINYRNLYIGEKFIIRR 59


RESULT 11
Q9QZD2
ID   Q9QZD2      PRELIMINARY;      PRT;   226 AA.
```

```
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
SQ   SEQUENCE   226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

  Query Match             2.5%;  Score 12;  DB 11;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 0.0011;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db         17 HHHHHHHHHHSS 28


RESULT 12
O42186
ID   O42186      PRELIMINARY;      PRT;   516 AA.
AC   O42186;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FTZ-F1 homolog.
GN   NR5A2 OR ZFF1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322094; PubMed=9178748;
RA   Liu D., Le Drean Y., Ekker M., Xiong F., Hew C.L.;
RT   "Teleost FTZ-F1 homolog and its splicing variant determine the
RT   expression of the salmon gonadotropin IIbeta subunit gene.";
RL   Mol. Endocrinol. 11:877-890(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AF014926; AAC60274.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04065; -.
DR   ZFIN; ZDB-GENE-990415-79; nr5a2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   516 AA;  58805 MW;  500DA09E116D5FA5 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 0.0023;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 13
Q90YL6
ID   Q90YL6      PRELIMINARY;      PRT;   517 AA.
AC   Q90YL6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nuclear receptor FF1aE.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chai C., Tan J.H., Gao W., Teh H.L., Chan W.K.;
RT   "Full length zebrafish ff1aE cDNA.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AY034061; AAK54449.1; -.
DR   HSSP; P19793; 1LBD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   517 AA;  58761 MW;  CF48D08AB7A75F29 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.0023;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 14
Q802D2
ID   Q802D2      PRELIMINARY;      PRT;   569 AA.
AC   Q802D2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
```

```
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048886; AAH48886.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
KW   Cyclin.
SQ   SEQUENCE   569 AA;  64117 MW;  812A7332C06EF4C2 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 0.0025;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        486 HHHHHHHHHHSS 497


RESULT 15
Q803K9
ID   Q803K9      PRELIMINARY;      PRT;   679 AA.
AC   Q803K9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T1 (Fragment).
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044435; AAH44435.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
KW   Cyclin.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  74887 MW;  AB5667D59439659F CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 679;
  Best Local Similarity   100.0%;  Pred. No. 0.0029;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        488 HHHHHHHHHHSS 499


Search completed: September 10, 2004, 05:12:45
Job time : 22.1967 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 31.1347 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 2540
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2540	100.0	472	2	AAW68393	Aaw68393 Clostridi
2	2307	90.8	472	2	AAW68394	Aaw68394 Clostridi
3	2144	84.4	1070	3	AAY93308	Aay93308 A mangane
4	2144	84.4	1095	3	AAY93311	Aay93311 A mangane
5	2137	84.1	1291	2	AAW68392	Aaw68392 Clostridi
6	2102	82.8	440	3	AAY77135	Aay77135 Synthetic
7	2102	82.8	440	4	AAB04091	Aab04091 Botulism
8	2084	82.0	439	4	AAB04085	Aab04085 Botulism
9	2060	81.1	848	4	AAB04082	Aab04082 Botulism
10	1586	62.4	1169	2	AAW56017	Aaw56017 Recombina
11	1350.5	53.2	473	2	AAW68400	Aaw68400 Clostridi
12	1191.5	46.9	449	3	AAY77139	Aay77139 Synthetic
13	1191.5	46.9	449	4	AAB04167	Aab04167 Botulism
14	992.5	39.1	462	2	AAR95009	Aar95009 Type A ne
15	992.5	39.1	462	2	AAW68390	Aaw68390 Clostridi
16	899	35.4	448	2	AAW68399	Aaw68399 Clostridi
17	876.5	34.5	452	2	AAW68396	Aaw68396 Clostridi
18	873	34.4	445	2	AAW68391	Aaw68391 Clostridi
19	858	33.8	451	2	AAW68395	Aaw68395 Clostridi
20	842.5	33.2	437	4	AAB04088	Aab04088 Botulism
21	842.5	33.2	438	2	AAR95008	Aar95008 Type A ne
22	842.5	33.2	438	2	AAW68389	Aaw68389 Clostridi
23	842.5	33.2	438	3	AAY77134	Aay77134 Synthetic
24	842.5	33.2	1067	3	AAY93307	Aay93307 A mangane
25	842.5	33.2	1092	3	AAY93310	Aay93310 A mangane
26	842.5	33.2	1296	2	AAR95010	Aar95010 C. botuli
27	839	33.0	435	4	AAB04090	Aab04090 Botulism
28	837.5	33.0	434	4	AAB04089	Aab04089 Botulism
29	829.5	32.7	1295	5	AAU99339	Aau99339 Clostridi
30	828.5	32.6	432	3	AAY77142	Aay77142 Native bo
31	828.5	32.6	837	3	AAY77140	Aay77140 Native bo
32	821.5	32.3	847	4	AAB04081	Aab04081 Botulism
33	790.5	31.1	415	4	AAB04083	Aab04083 Botulism
34	777	30.6	685	4	AAE07893	Aae07893 Modified
35	774.5	30.5	862	4	AAE07890	Aae07890 Modified
36	774.5	30.5	887	4	AAE07892	Aae07892 Modified
37	774.5	30.5	979	6	AAE35713	Aae35713 BoNT/F-Hc
38	774.5	30.5	1192	6	AAE35711	Aae35711 BoNT/F-Hc
39	774.5	30.5	1192	6	AAE35710	Aae35710 BoNT/F-Hc
40	774	30.5	1092	4	AAE07900	Aae07900 C. botuli
41	773.5	30.5	645	4	AAE07894	Aae07894 Modified
42	773.5	30.5	645	6	AAE35692	Aae35692 DipT HN d
43	773.5	30.5	657	6	AAE35693	Aae35693 BoNT/F-Hc
44	773.5	30.5	657	6	AAE35694	Aae35694 BoNT/F-Hc
45	773	30.4	432	4	AAB04103	Aab04103 Botulism

ALIGNMENTS

RESULT 1
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant

```
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             100.0%;  Score 2540;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-178;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 2
AAW68394
ID   AAW68394 standard; protein; 472 AA.
XX
AC   AAW68394;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30581.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 303-305; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             90.8%;  Score 2307;  DB 2;  Length 472;
  Best Local Similarity   89.8%;  Pred. No. 4.7e-161;
  Matches  424;  Conservative   26;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              |||||||||||:|||||||||||||:|||||||||||||||:|||||||||||||||:|
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||:|||||||| |||| ||||| ||| |||| |||||:||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              |||||:|:||||:| ||||||||||||||||||||||||||||||||||| ||| ||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||| |||||||| |||||:||||||||||:||||||||||||||||:||    ::|||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              || |||||::||||||: ||||:||| ||:||||||||:|||||||||||||||:|||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||:: ::|||||||||||||||||||   |||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 3
AAY93308
ID   AAY93308 standard; protein; 1070 AA.
XX
AC   AAY93308;
XX
DT   04-SEP-2000  (first entry)
XX
```

DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype B.
XX
OS Synthetic.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 45-47; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype B. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a domain
CC that translocates the SOD of the composition into the neuronal cell.
CC After translocation, the linker is cleaved to release the SOD. The
CC composition is useful for treating neuronal diseases caused or augmented
CC by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC disease, Huntington's disease and motor neurone diseases
XX
SQ Sequence 1070 AA;

Query Match 84.4%; Score 2144; DB 3; Length 1070;
Best Local Similarity 89.2%; Pred. No. 1.3e-148;
Matches 398; Conservative 26; Mismatches 22; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              ||||||||||||||||||||||||:|||||||||||||||||||:||||||||||||||
Db        625 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 684

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|||||||||||||||:||||||||||||||:|| ||||||||||||||||||||||||
Db        685 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 744

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||||| ||||
Db        745 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 804

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              ||||| ||| |||| ||||||:||||||||||||||:|:|||||:| ||||||||||||
Db        805 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 864

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              ||||||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db        865 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 924

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||||:||    ::||||| |||||::||||||: ||||:|||
Db        925 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 984

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              ||:||||||||:|||||||||||||||:|||||||||||||:: ::||||||||||||||
Db        985 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1044

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||||||   ||||||||||||||||
Db       1045 EVKRKPYNLKLGCNWQFIPKDEGWTE 1070
```

RESULT 4
AAY93311
ID AAY93311 standard; protein; 1095 AA.
XX
AC AAY93311;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype B.
XX
OS Synthetic.
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 54-57; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide dismutase
CC (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC serotype B. The specification describes a composition for delivery of SOD
CC to neuronal cells. The composition comprises SOD linked, by a cleavable
CC linker, to a neuronal cell targeting component (NCTC). This component has
CC a domain that binds to a neuronal cell and a domain that translocates the
CC SOD of the composition into the neuronal cell. After translocation, the
CC linker is cleaved to release the SOD. The composition is useful for
CC treating neuronal diseases caused or augmented by oxidative stress, such
CC as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases
XX
SQ Sequence 1095 AA;

Query Match 84.4%; Score 2144; DB 3; Length 1095;
Best Local Similarity 89.2%; Pred. No. 1.3e-148;
Matches 398; Conservative 26; Mismatches 22; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
```

```
                  |||||||||||||||||||||||||:|||||||||||||||||||:|||||||||||||
Db        650 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 709

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :||||||||||||||:|||||||||||||||:|| ||||||||||||||||||||||||
Db        710 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 769

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||||||||||||||||||||||||||||||||||||||||||||:||||||| ||||
Db        770 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 829

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
               ||||| ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||
Db        830 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 889

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db        890 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 949

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    ::||||| |||||::||||||: ||||:|||
Db        950 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1009

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
               ||:|||||||:|||||||||||||:|||||||||:: ::|||||||||||||
Db       1010 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1069

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||||||   ||||||||||||||||
Db       1070 EVKRKPYNLKLGCNWQFIPKDEGWTE 1095
```

RESULT 5
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30579.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 291-296; 428pp; English.
XX
CC This is the amino acid sequence of the type B toxin of Clostridium
CC botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC B toxin has been expressed as a histidine-tagged protein in Escherichia
CC coli. The invention relates to C. botulinum recombinant toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of the recombinant proteins are E. coli, insect cells and
CC yeast cells. The recombinant proteins are used as immunogens for the
CC production of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1291 AA;

Query Match 84.1%; Score 2137; DB 2; Length 1291;
Best Local Similarity 89.0%; Pred. No. 5.3e-148;
Matches 397; Conservative 26; Mismatches 23; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              |||||||||||||||||||||||:|||||||||||||||||||:||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :||||||||||||||:|||||||||||||||:|| ||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||||||||||||||||||||||||||||||||||||||||||||:||||||| ||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
               ||||| ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    ::||||| |||||::||||||: ||||:|||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
               ||:|||| |||:||||||||||||:|||||||||:: ::|||||||||||||
Db       1206 NTIQIKEYHEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||||||   ||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 6
AAY77135
ID AAY77135 standard; protein; 440 AA.
XX
AC AAY77135;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype B (BoNTB) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype B; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX

```
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87213.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 23; Page; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast. Note: This
CC   sequence is not given in the specification, but is decoded from the BoNTB
CC   Hc DNA sequence given on pages 39-40
XX
SQ   Sequence 440 AA;

  Query Match             82.8%;  Score 2102;  DB 3;  Length 440;
  Best Local Similarity   88.6%;  Pred. No. 4.9e-146;
  Matches  390;  Conservative   26;  Mismatches   24;  Indels    0;  Gaps    0;

Qy         33 MFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRV 92
              | |||||||||||||||||:|||||||||||||||||||:|||||||||||||:|||||
Db          1 MANKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 60

Qy         93 TQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRI 152
              |||||||||:|||||||||||||:|| |||||||||||||||||||||||||||||||
Db         61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy        153 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDI 212
              |||||||||||||||||||||||||||||||||||||||||:||||||| |||| |||||
Db        121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 180

Qy        213 GEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNP 272
               ||| |||| |||||:||||||||||||||:|:||||:| ||||||||||||||||||
Db        181 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 240

Qy        273 LMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRES 332
              |||||||||||||||||| ||| ||||| |||||||| |||||:||||||||||:|
Db        241 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 300

Qy        333 NSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIK 392
              |||||||||||||||:||    ::||||| |||||::||||||: ||||:| | ||:||
Db        301 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDELYNTIQIK 360

Qy        393 EYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKP 452
              ||||||:|||||||||||||:|||||||||||:: ::||||||||||||||||||||
Db        361 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 420

Qy        453 YKSNLGCNWQFIPKDEGWTE 472
              |   ||||||||||||||||
Db        421 YNLKLGCNWQFIPKDEGWTE 440


RESULT 7
AAB04091
ID   AAB04091 standard; protein; 440 AA.
XX
AC   AAB04091;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype B).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54485.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 4b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 440 AA;

  Query Match             82.8%;  Score 2102;  DB 4;  Length 440;
  Best Local Similarity   88.6%;  Pred. No. 4.9e-146;
  Matches  390;  Conservative   26;  Mismatches   24;  Indels    0;  Gaps    0;

Qy         33 MFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRV 92
              | |||||||||||||||||:|||||||||||||||||||:|||||||||||||:|||||
Db          1 MANKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 60
```

```
Qy          93 TQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRI 152
               ||||||||||:|||||||||||||||:|| |||||||||||||||||||||||||||||||
Db          61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy         153 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDI 212
               ||||||||||||||||||||||||||||||||||||||||:|||||||| |||| |||||
Db         121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 180

Qy         213 GEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNP 272
               ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||||||||||
Db         181 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 240

Qy         273 LMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRES 332
               |||||||||||||||||||| ||| ||||| |||||||| |||||:||||||||||:|
Db         241 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 300

Qy         333 NSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIK 392
               |||||||||||||||:||    ::||||| |||||::|||||: |||||:| | ||:||
Db         301 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDELYNTIQIK 360

Qy         393 EYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKP 452
               ||||||:|||||||||||||:|||||||||||||:: ::||||||||||||||||||||
Db         361 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 420

Qy         453 YKSNLGCNWQFIPKDEGWTE 472
               |   ||||||||||||||||
Db         421 YNLKLGCNWQFIPKDEGWTE 440
```

RESULT 8
AAB04085
ID AAB04085 standard; protein; 439 AA.
XX
AC AAB04085;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin C fragment sequence (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Example 8; Page 38; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 439 AA;

Query Match 82.0%; Score 2084; DB 4; Length 439;
Best Local Similarity 88.4%; Pred. No. 1e-144;
Matches 388; Conservative 26; Mismatches 25; Indels 0; Gaps 0;

```
Qy          34 FNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVT 93
               ||||||||||||||||:|||||||||||||||||||:|||||||||||||:||||||
Db           1 FNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVT 60

Qy          94 QNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
               |||||||||:|||||||||||||:|| |||||||||||||||||||||||||||||||
Db          61 QNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIT 120

Qy         154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDIG 213
               ||||||||||||||||||||||||||||||||||||||:|||||||| |||| |||||
Db         121 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIR 180

Qy         214 EVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPL 273
               ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||||||||||
Db         181 EVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPL 240

Qy         274 MYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
               |||||||||||||||||||| ||| ||||| |||||||| |||||:||||||||||:||
Db         241 MYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSN 300

Qy         334 SQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKE 393
               |||||||||||||||:||    ::| ||| || ||::||||||: ||||:||| ||:|||
Db         301 SQSINDDIVRKEDYIYLDFFNLNQEQRVYTYKKFKKEEEKLFLAPISDSDEFYNTIQIKE 360

Qy         394 YDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPY 453
               |||||:|||||||||||||:|||||||||||||:: ::||||||||||||||||||||
Db         361 YDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPY 420

Qy         454 KSNLGCNWQFIPKDEGWTE 472
                  ||||||||||||||||
Db         421 NLKLGCNWQFIPKDEGWTE 439
```

RESULT 9
AAB04082
ID AAB04082 standard; protein; 848 AA.
XX
AC AAB04082;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX

```
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Disclosure; Page 13; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 848 AA;

  Query Match             81.1%;  Score 2060;  DB 4;  Length 848;
  Best Local Similarity   87.5%;  Pred. No. 1.4e-142;
  Matches  391;  Conservative   26;  Mismatches   26;  Indels    4;  Gaps    4;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              |||||||||||||||||||||||:||||||||||||||||||:||||||||||||||
Db        405 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 464

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|||| |||||||||:|||||||||||||:|| ||||||||||||||||||||||||||
Db        465 NSKIRTTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 524

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              ||| |||||||||||||||||||||||||||||||||||||:|||||||| ||||
Db        525 IRG-RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 583

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              ||||| ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||
Db        584 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 643

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSY-IKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEK 325
              |||||||||||||||||||||| ||| ||| ||||| |||||||| |||||:||||||
Db        644 DFWGNPLMYNKEYYMFNAGNKNSYTIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 703

Qy        326 FIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEF 385
              |||||:|||||||||||||||:||    ::||||| |||||::|| |||: ||||:||
Db        704 FIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEE-LFLAPISDSDEF 762

Qy        386 YKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYL 445
              | ||:||||||||:|||||||||||||||:||||||||||:: :: || |||| |||
Db        763 YNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEVKDVFCISDWYL 822

Qy        446 KEVKRKPYKSNLGCNWQFIPKDEGWTE 472
```

```
              |||||||   ||||||||||||||||
Db        823 -EVKRKPYNLKLGCNWQFIPKDEGWTE 848


RESULT 10
AAW56017
ID   AAW56017 standard; protein; 1169 AA.
XX
AC   AAW56017;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH728/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW   detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB002273.
XX
PR   23-AUG-1996;   96GB-00017671.
PR   13-DEC-1996;   96GB-00025996.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Shone CC,  Quinn CP,  Foster KA;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26289.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT   immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS   Example 2; Page 91-94; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from the
CC   present invention. The present invention describes recombinant neurotoxin
CC   proteins which comprise a first and second domain, where the first domain
CC   is adapted to cleave one or more vesicle or plasma-membrane associated
CC   proteins essential to exocytosis, and where the second domain is adapted:
CC   (a) to translocate the protein into a cell; (b) to increase the
CC   solubility of the protein compared to the solubility of the first domain
CC   on its own, or (c) both to translocate the protein into a cell and to
CC   increase the solubility of the protein compared to the solubility of the
CC   first domain on its own, the protein being free of clostridial neurotoxin
CC   (CN) and free of CN precursor that can be converted into toxin by
CC   proteolytic action. The recombinant proteins can be used as therapeutic
CC   agents for targeting cells expressing a relevant substrate. The products
CC   can also be used as immunogens and as non-toxic standards for the
CC   assessment and development of in vitro assays for the detection of
CC   functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC   environmental samples
XX
SQ   Sequence 1169 AA;

  Query Match             62.4%;  Score 1586;  DB 2;  Length 1169;
  Best Local Similarity   92.6%;  Pred. No. 1.4e-107;
  Matches  300;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              |||||||||||||||||||||||:||||||||||||||||||:||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|||||||||||||:|||||||||||||:|| ||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965
```

```
Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||| ||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              ||||| ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHL 350
              ||||:||||||||||||||||:|
Db       1146 IIRRKSNSQSINDDIVRKEDYIYL 1169
```

RESULT 11
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30596.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC express BotG soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX

SQ Sequence 473 AA;

Query Match 53.2%; Score 1350.5; DB 2; Length 473;
Best Local Similarity 52.9%; Pred. No. 8.5e-91;
Matches 253; Conservative 94; Mismatches 120; Indels 11; Gaps 7;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||::|| | | | :| ||:| ||   ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGAKVEVYDGVKLND--KNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKY 118
              |||| : |   | ||    |||| :| :| |   |:: ::::||| :||::||:| |||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        119 RNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178
               |:||| |: |||||:|:||:||||:||:||||||||:| |:||:||||:|:::||:
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        179 YINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWM 237
              |||:|| :||||: | || |||||:|: :  | ::  :  : :| |||    | |:|:|:
Db        181 YINKWFSITITNDRLGNANIYIYNGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        238 KYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDS 297
              | |:||  :|| : :  :| ||| :  |||||||| |: :||:|| | :| |||    :
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        298 SVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLH 355
              |:||   |: :|  :  |||:|||:| :|||:: |||::| ||:|||: |||:|::  :
Db        301 SMGETAPRTNFNNAA--INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNIS 358

Qy        356 HEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDD 415
              | :|||     || : :|||: |:|   ||  ::||:| |: :|:||:| :||    |
Db        359 DESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKT 415

Qy        416 IGLIGIHRFYES-GVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              || || :| :  | :  | :|||||:|||: :     |  |||||||| ||||||
Db        416 FGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
```

RESULT 12
AAY77139
ID AAY77139 standard; protein; 449 AA.
XX
AC AAY77139;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype G (BoNTG) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype G; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87217.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum

PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 28; Page 47-48; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 449 AA;

Query Match 46.9%; Score 1191.5; DB 3; Length 449;
Best Local Similarity 50.2%; Pred. No. 3.8e-79;
Matches 228; Conservative 94; Mismatches 121; Indels 11; Gaps 7;

```
Qy         25 MADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKL 82
              | |||||::|| | | | :| ||:| ||   ||| ||||| : |   |  ||    ||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         83 TSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 142
               :| :| |   |:: ::::||| :||::||:| ||| |:||| |: ||||||:|:||:||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 201
              ||:||:||||||||||:| |:||:||||:|:::||:|||:|| :|||: | || |||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        202 TLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSY 261
              :|: :  | ::  :  : :| |||    | |:|:|:| |:||  :|| : :  :| |||
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        262 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLY 321
              :  |||||||||| |: :||:|| | :| |||    :|:||   |: :|  :  |||:|||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLY 298

Qy        322 IGEKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQEEKLFLSII 379
              :| :|||:: |||::| ||:|||: |||:|::  :  | :|||     || : :|||: |
Db        299 LGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPI 358

Qy        380 SDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GVLRKKYKDYF 438
              :|   ||  ::||:| |: :|:||:| :||   |   || || :| :  | :   | :||
Db        359 NDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYF 415

Qy        439 CISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||:|||: :     |  |||||||| ||||||
Db        416 CISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 449
```

RESULT 13
AAB04167
ID AAB04167 standard; protein; 449 AA.
XX
AC AAB04167;
XX
DT 11-APR-2001 (first entry)

XX
DE Botulism toxin heavy chain C-terminal sequence (serotype G).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54491.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 10b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 449 AA;

Query Match 46.9%; Score 1191.5; DB 4; Length 449;
Best Local Similarity 50.2%; Pred. No. 3.8e-79;
Matches 228; Conservative 94; Mismatches 121; Indels 11; Gaps 7;

```
Qy         25 MADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKL 82
              | |||||::|| | | | :| ||:| ||   ||| ||||| : |   |  ||    ||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         83 TSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 142
               :| :| |   |:: ::::||| :||::||:| ||| |:||| |: ||||||:|:||:||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 201
              ||:||:||||||||||:| |:||:||||:|:::||:|||:|| :|||: | || |||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        202 TLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSY 261
              :|: :  | ::  :  : :| |||    | |:|:|:| |:||  :|| : :  :| |||
```

```
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        262 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLY 321
              :  ||||||||| |: :||:|| | :| |||    :|:||   |: :|  :  |||:|||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLY 298

Qy        322 IGEKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQEEKLFLSII 379
              :| :|||:: |||::| ||:|||: |||:|:: :  | :|||     || : :|||: |
Db        299 LGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYYVLVNSKEIQTQLFLAPI 358

Qy        380 SDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GVLRKKYKDYF 438
              :|   ||  ::||:| |: :|:||:| :||   |   || || :| :  | :   | :||
Db        359 NDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYF 415

Qy        439 CISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||:|||: :     |  |||||||| ||||||
Db        416 CISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 14
AAR95009
ID   AAR95009 standard; protein; 462 AA.
XX
AC   AAR95009;
XX
DT   07-JUL-1996  (first entry)
XX
DE   Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW   Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW   Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= Polyhistidine_tag
FT   Protein         22. .462
FT                   /label= C-fragment
XX
PN   WO9612802-A1.
XX
PD   02-MAY-1996.
XX
PF   23-OCT-1995;   95WO-US013737.
XX
PR   24-OCT-1994;   94US-00329154.
PR   16-MAR-1995;   95US-00405496.
PR   14-APR-1995;   95US-00422711.
PR   07-JUN-1995;   95US-00480604.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Padhye NV,  Kink JA,  Thalley BS,  Stafford DC;
PI   Firca JR;
XX
DR   WPI; 1996-230603/23.
DR   N-PSDB; AAT29246.
XX
PT   Fusion proteins comprising non-toxin protein and part of toxin - useful
PT   to form anti-toxins against Clostridium botulinum type A, and C.
PT   difficile type toxins, and to treat C. difficile intoxication, partic.
PT   diarrhoea.
XX
PS   Claim 7; Page 340-342; 434pp; English.
XX
CC   pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC   (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC   fragment C (see also AAR95008) of the Clostridium botulinum type A
CC   neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC   soluble protein and was purified by metal chelate affinity chromatography
CC   to obtain a product free of endotoxin contamination that may be useful as
CC   an immunogen in vaccine compsns
XX
SQ   Sequence 462 AA;

  Query Match             39.1%;  Score 992.5;  DB 2;  Length 462;
  Best Local Similarity   43.6%;  Pred. No. 1.6e-64;
  Matches  210;  Conservative   86;  Mismatches  153;  Indels   33;  Gaps   14;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              |||||||||||||||||||||||||    |:  | :|   |:|  ||||||  |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVKLN----DKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116
                | :|: :  | |:|    ||||  :| :   |||  |     |::|||: :|| |||||||
Db         59 RYASKINI--GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIP 116

Qy        117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176
              || |    |   |||||||||:||||||:|:    |||||  |     : | |:|:   :|
Db        117 KYFNSISLN---NEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINI 173

Qy        177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFI 235
              |:|||| ||||||| |:|:|||||| |     | ::| :  :  | |||||  |  ::|
Db        174 SDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYI 233

Qy        236 WMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVK 295
              |:|||::|: :||:  ||::|  || |  ||||||: | |:| ||| |  : | |: :
Db        234 WIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNN 293

Qy        296 DSSVGEILI---RSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDL 352
                  | : :   |       : |:|  :|| | ||||:: ::      |:|||  | :::::
Db        294 VGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGN--KDNIVRNNDRVYINV 350

Qy        353 VLHHEEWRVYAYKYFKEQEEKLFLSI-ISDSNEFYKTIEIKEYDEQP-SYSCQLLFKKDE 410
              |: ::|:|: |    :   ||:  :: | |     : : :|  ::|  :  |::     :
Db        351 VVKNKEYRL-ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQD 407

Qy        411 ESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGW 470
              : :||| || |:|     |          | || ::::|     |||:|:||| |:||
Db        408 NNGNDIGFIGFHQFNNIAKL--------VASNWYNRQIERS--SRTLGCSWEFIPVDDGW 457

Qy        471 TE 472
              ||
Db        458 GE 459


RESULT 15
AAW68390
ID   AAW68390 standard; protein; 462 AA.
XX
AC   AAW68390;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW   Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW   botulism.
XX
OS   Clostridium botulinum; serotype A.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal histidine tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
```

XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30572, AAV30575.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 24; Page 265-267; 428pp; English.
XX
CC This is the amino acid sequence of a histidine-tagged fragment C
CC polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on a poly-histidine binding
CC affinity resin. The invention relates to recombinant proteins derived
CC from C. botulinum toxins. Methods are provided which allow the isolation
CC of soluble recombinant proteins that are free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant proteins are
CC E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are useful
CC in the treatment of humans and animals at risk of intoxication with
CC clostridial toxin
XX
SQ Sequence 462 AA;

Query Match 39.1%; Score 992.5; DB 2; Length 462;
Best Local Similarity 43.6%; Pred. No. 1.6e-64;
Matches 210; Conservative 86; Mismatches 153; Indels 33; Gaps 14;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||   |:  | :|   |:|  ||||||  |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVKLN----DKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116
               | :|: :  | |:|    |||| :| :   ||| |     |::|||: :|| |||||||
Db         59 RYASKINI--GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIP 116

Qy        117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176
              || |    |   |||||||||:||||||:|:    ||||| |     : | |:|:   :|
Db        117 KYFNSISLN---NEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINI 173

Qy        177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFI 235
              |:|||| ||||||| |:|:|||||| |      | ::| :  :  | |||||  |  ::|
Db        174 SDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYI 233

Qy        236 WMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVK 295
              |:|||::|: :||:  ||::|  || |  ||||||: | |:| |||  |  : | |: :
Db        234 WIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNN 293

Qy        296 DSSVGEILI---RSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDL 352
                | : :   |      : |:|  :|| | ||||:: ::     |:|||  | :::::
Db        294 VGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGN--KDNIVRNNDRVYINV 350

Qy        353 VLHHEEWRVYAYKYFKEQEEKLFLSI-ISDSNEFYKTIEIKEYDEQP-SYSCQLLFKKDE 410
              |: ::|:|: |    :   ||:  :: | |     : : :|  ::|  :  |::     :
Db        351 VVKNKEYRL-ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQD 407

Qy        411 ESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGW 470
               : :||| || |:|     |           | || ::::|      |||:|:||| |:||
Db        408 NNGNDIGFIGFHQFNNIAKL--------VASNWYNRQIERS--SRTLGCSWEFIPVDDGW 457

Qy        471 TE 472
               |
Db        458 GE 459
```

Search completed: September 10, 2004, 03:33:25
Job time : 34.1347 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 10.0183 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 2540
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1586	62.4	1169	4	US-09-255-829-20	Sequence 20, Appl
2	992.5	39.1	462	1	US-08-480-604A-26	Sequence 26, Appl
3	992.5	39.1	462	2	US-08-405-496A-26	Sequence 26, Appl
4	992.5	39.1	462	3	US-08-915-136-26	Sequence 26, Appl
5	992.5	39.1	462	4	US-09-084-517-26	Sequence 26, Appl
6	842.5	33.2	438	1	US-08-480-604A-23	Sequence 23, Appl
7	842.5	33.2	438	2	US-08-405-496A-23	Sequence 23, Appl
8	842.5	33.2	438	3	US-08-915-136-23	Sequence 23, Appl
9	842.5	33.2	438	4	US-09-084-517-23	Sequence 23, Appl
10	842.5	33.2	1296	1	US-08-480-604A-28	Sequence 28, Appl
11	842.5	33.2	1296	2	US-08-405-496A-28	Sequence 28, Appl
12	842.5	33.2	1296	3	US-08-915-136-28	Sequence 28, Appl
13	842.5	33.2	1296	4	US-09-084-517-28	Sequence 28, Appl
14	608	23.9	452	1	US-07-618-312A-2	Sequence 2, Appli
15	608	23.9	452	1	US-08-280-228-2	Sequence 2, Appli
16	607	23.9	452	1	US-07-618-312A-4	Sequence 4, Appli
17	607	23.9	452	1	US-08-280-228-4	Sequence 4, Appli
18	607	23.9	618	1	US-08-668-381A-5	Sequence 5, Appli
19	607	23.9	853	4	US-08-913-880C-17	Sequence 17, Appl
20	607	23.9	858	4	US-08-913-880C-16	Sequence 16, Appl
21	607	23.9	860	4	US-08-913-880C-15	Sequence 15, Appl
22	607	23.9	862	4	US-08-913-880C-14	Sequence 14, Appl
23	607	23.9	865	4	US-08-913-880C-13	Sequence 13, Appl
24	607	23.9	866	4	US-08-913-880C-12	Sequence 12, Appl
25	607	23.9	874	4	US-08-913-880C-11	Sequence 11, Appl
26	607	23.9	875	4	US-08-913-880C-10	Sequence 10, Appl
27	607	23.9	1315	4	US-08-913-880C-1	Sequence 1, Appli
28	588	23.1	452	1	US-08-110-786A-8	Sequence 8, Appli
29	175.5	6.9	141	4	US-09-465-276-1	Sequence 1, Appli
30	169.5	6.7	140	3	US-08-446-114A-22	Sequence 22, Appl
31	168.5	6.6	803	4	US-09-543-681A-4886	Sequence 4886, Ap
32	152	6.0	2391	2	US-08-446-855A-2	Sequence 2, Appli
33	152	6.0	2391	3	US-09-150-741-2	Sequence 2, Appli
34	150.5	5.9	2184	4	US-09-417-485D-6	Sequence 6, Appli
35	150	5.9	23	1	US-08-480-604A-24	Sequence 24, Appl
36	150	5.9	23	2	US-08-405-496A-24	Sequence 24, Appl
37	150	5.9	23	3	US-08-915-136-24	Sequence 24, Appl
38	150	5.9	23	4	US-09-084-517-24	Sequence 24, Appl
39	150	5.9	794	4	US-09-417-485D-8	Sequence 8, Appli
40	147	5.8	323	1	US-08-591-989-5	Sequence 5, Appli
41	146	5.7	22	4	US-09-660-742-3	Sequence 3, Appli
42	146	5.7	53	2	US-08-651-818A-19	Sequence 19, Appl
43	146	5.7	53	3	US-09-184-826-19	Sequence 19, Appl
44	146	5.7	54	2	US-08-651-818A-23	Sequence 23, Appl
45	146	5.7	54	3	US-09-184-826-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 62.4%; Score 1586; DB 4; Length 1169;
Best Local Similarity 92.6%; Pred. No. 8.7e-114;
Matches 300; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

```
Qy          27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
               ||||||||||||||||||||||||:||||||||||||||||||:|||||||||||||
Db         846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy          87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
               :|||||||||||||:|||||||||||||:|| ||||||||||||||||||||||||||
Db         906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy         147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
               ||||||||||||||||||||||||||||||||||||||||||:|||||||| ||||
Db         966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy         207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
               ||||| ||| |||| ||||||:||||||||||||||:|:||||:| ||||||||||
Db        1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy         267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
               |||||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db        1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy         327 IIRRESNSQSINDDIVRKEDYIHL 350
               ||||:||||||||||||||:|
Db        1146 IIRRKSNSQSINDDIVRKEDYIYL 1169


RESULT 2
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match           39.1%;  Score 992.5;  DB 1;  Length 462;
  Best Local Similarity  43.6%;  Pred. No. 1.3e-68;
  Matches  210;  Conservative   86;  Mismatches  153;  Indels   33;  Gaps   14;

Qy           1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
               |||||||||||||||||||||||||||   |:  | :|   |:|  ||||||  |:|||||
Db           1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy          61 GYGAKVEVYDGVKLN----DKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116
                 | :|: :  | |:|    |||| :| :   ||| |      |::|||: :|| |||||||
Db          59 RYASKINI--GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIP 116

Qy         117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176
               || |    |  |||||||||:||||||:|:    ||||| |     : | |:|:   :|
Db         117 KYFNSISLN---NEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINI 173

Qy         177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFI 235
               |:|||| ||||||| |:|:|||||| |     | ::| :  :  | |||||  |  ::|
Db         174 SDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYI 233

Qy         236 WMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVK 295
               |:|||::|: :||:  ||::|  || |  ||||||: | |:| ||| |  : | |: :
Db         234 WIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNN 293

Qy         296 DSSVGEILI---RSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDL 352
                 | : :   |     : |:|  :|| | ||||:: ::     |:|||  | ::::::
Db         294 VGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGN--KDNIVRNNDRVYINV 350

Qy         353 VLHHEEWRVYAYKYFKEQEEKLFLSI-ISDSNEFYKTIEIKEYDEQP-SYSCQLLFKKDE 410
               |: ::|:|: |    :   ||:  :: | |     : : :|  ::|  :  |::      :
Db         351 VVKNKEYRL-ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQD 407

Qy         411 ESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGW 470
                : :||| || |:|     |          | || ::::|      |||:|:||| |:||
Db         408 NNGNDIGFIGFHQFNNIAKL--------VASNWYNRQIERS--SRTLGCSWEFIPVDDGW 457

Qy         471 TE 472
               ||
Db         458 GE 459


RESULT 3
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/405,496A
;       FILING DATE:  16-MAR-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  INGOLIA, DIANE E.
;       REGISTRATION NUMBER:  40,027
;       REFERENCE/DOCKET NUMBER:  OPHD-01308
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  462 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match           39.1%;  Score 992.5;  DB 2;  Length 462;
  Best Local Similarity  43.6%;  Pred. No. 1.3e-68;
  Matches  210;  Conservative   86;  Mismatches  153;  Indels   33;  Gaps   14;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              |||||||||||||||||||||||||   |:  | :|   |:|  ||||||  |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVKLN----DKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116
                | :|: :  | |:|    ||||  :| :   ||| |     |::|||: :|| |||||||
Db         59 RYASKINI--GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIP 116

Qy        117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176
              || |    |   |||||||||:||||||:|:    |||||  |     : | |:|:   :|
Db        117 KYFNSISLN---NEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINI 173

Qy        177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFI 235
              |:|||| ||||||| |:|:|||||| |      | ::| :  :  | |||||  |  ::|
Db        174 SDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYI 233

Qy        236 WMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVK 295
              |:|||::|: :||:  ||::|  || |  ||||||: | |:| ||| |  : | |: :
Db        234 WIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNN 293

Qy        296 DSSVGEILI---RSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDL 352
                  | : :   |      : |:|  :|| | ||||:: ::      |:|||  | :::::
Db        294 VGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGN--KDNIVRNNDRVYINV 350

Qy        353 VLHHEEWRVYAYKYFKEQEEKLFLSI-ISDSNEFYKTIEIKEYDEQP-SYSCQLLFKKDE 410
              |: ::|:|: |    :   ||:  :: | |     : : :|  ::|  :  |::     :
Db        351 VVKNKEYRL-ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQD 407

Qy        411 ESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGW 470
              : :||| || |:|     |          | || ::::|      |||:|:||| |:||
Db        408 NNGNDIGFIGFHQFNNIAKL--------VASNWYNRQIERS--SRTLGCSWEFIPVDDGW 457

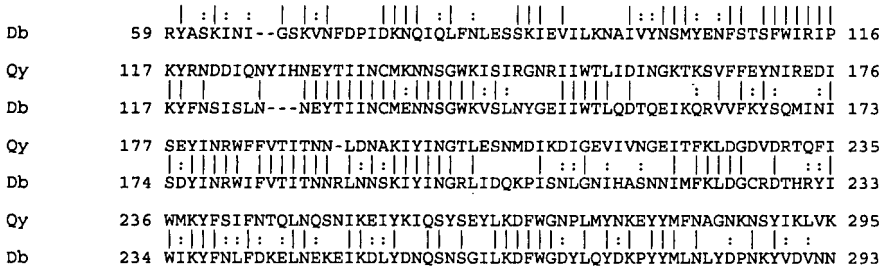
Qy        471 TE 472
              |
Db        458 GE 459
```

```
RESULT 4
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match           39.1%;  Score 992.5;  DB 3;  Length 462;
  Best Local Similarity  43.6%;  Pred. No. 1.3e-68;
  Matches  210;  Conservative   86;  Mismatches  153;  Indels   33;  Gaps   14;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              |||||||||||||||||||||||||   |:  | :|   |:|  ||||||  |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVKLN----DKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116
```



```
Db         59 RYASKINI--GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIP 116

Qy        117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176

Db        117 KYFNSISLN---NEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINI 173

Qy        177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFI 235

Db        174 SDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYI 233

Qy        236 WMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVK 295

Db        234 WIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNN 293

Qy        296 DSSVGEILI---RSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDL 352

Db        294 VGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGN--KDNIVRNNDRVYINV 350

Qy        353 VLHHEEWRVYAYKYFKEQEEKLFLSI-ISDSNEFYKTIEIKEYDEQP-SYSCQLLFKKDE 410

Db        351 VVKNKEYRL-ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQD 407

Qy        411 ESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGW 470

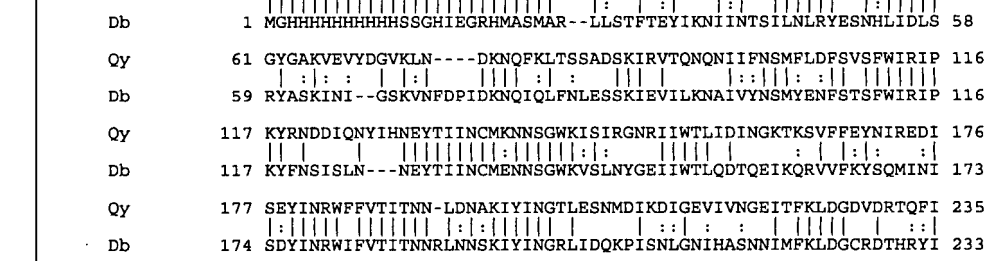
Db        408 NNGNDIGFIGFHQFNNIAKL--------VASNWYNRQIERS--SRTLGCSWEFIPVDDGW 457

Qy        471 TE 472

Db        458 GE 459
```

RESULT 5
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-26

Query Match 39.1%; Score 992.5; DB 4; Length 462;
Best Local Similarity 43.6%; Pred. No. 1.3e-68;
Matches 210; Conservative 86; Mismatches 153; Indels 33; Gaps 14;



```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVKLN----DKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116

Db         59 RYASKINI--GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIP 116

Qy        117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176

Db        117 KYFNSISLN---NEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINI 173

Qy        177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFI 235

Db        174 SDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYI 233

Qy        236 WMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVK 295

Db        234 WIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNN 293

Qy        296 DSSVGEILI---RSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDL 352

Db        294 VGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGN--KDNIVRNNDRVYINV 350

Qy        353 VLHHEEWRVYAYKYFKEQEEKLFLSI-ISDSNEFYKTIEIKEYDEQP-SYSCQLLFKKDE 410

Db        351 VVKNKEYRL-ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQD 407

Qy        411 ESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGW 470

Db        408 NNGNDIGFIGFHQFNNIAKL--------VASNWYNRQIERS--SRTLGCSWEFIPVDDGW 457

Qy        471 TE 472

Db        458 GE 459
```

RESULT 6
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO

```
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match            33.2%;  Score 842.5;  DB 1;  Length 438;
  Best Local Similarity  40.6%;  Pred. No. 4.4e-57;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:|||||  | :|: :  | |:|    |||| :| :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 61

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                 |||  |     |::|||: :|| |||||||| |    |   |||||||||:||||||:
Db         62 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 118

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:||||| ||||||| |:|:|||||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
                | ::| :  :  | |||||  |  ::||:|||::|: :||:  ||::|  || |
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :       | : :   |       : |:|  :||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
              | ||||:: ::     |:|||  | :::::|: ::|:|: |    :   ||:  :: |
Db        298 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 354

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
              |     : : :|  ::|  :  |::     : : :||| || |:|     |
Db        355 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 404

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
                | || ::::|      |||:|:||| |:|| |
Db        405 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 7
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match            33.2%;  Score 842.5;  DB 2;  Length 438;
  Best Local Similarity  40.6%;  Pred. No. 4.4e-57;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:|||||  | :|: :  | |:|    |||| :| :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 61

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                 |||  |     |::|||: :|| |||||||| |    |   |||||||||:||||||:
Db         62 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 118
```

```
Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:||||| ||||||| |:|:|||||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
              | ::| :  :  | ||||| |  ::||:|||::|: :||:  ||::|  || |
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :      | : :   |      : |:|  :||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
              | ||||:: ::     |:|||  | :::::|: ::|:|: |    :   ||:  :: |
Db        298 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 354

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
              |     : : :|  ::|  :  |::     : : :||| || |:|      |
Db        355 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 404

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              | || ::::|      |||:|:||| |:|| |
Db        405 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 8
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match             33.2%;  Score 842.5;  DB 3;  Length 438;
  Best Local Similarity   40.6%;  Pred. No. 4.4e-57;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:||||| | :|: :  | |:|    |||| :| :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 61

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                 ||| |     |::|||: :|| |||||||| |     |   |||||||||:||||||:
Db         62 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 118

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:||||| ||||||| |:|:|||||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
              | ::| :  :  | ||||| |  ::||:|||::|: :||:  ||::|  || |
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :      | : :   |      : |:|  :||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
              | ||||:: ::     |:|||  | :::::|: ::|:|: |    :   ||:  :: |
Db        298 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 354

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
              |     : : :|  ::|  :  |::     : : :||| || |:|      |
Db        355 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 404

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              | || ::::|      |||:|:||| |:|| |
Db        405 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 9
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-23

  Query Match             33.2%;  Score 842.5;  DB 4;  Length 438;
  Best Local Similarity   40.6%;  Pred. No. 4.4e-57;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:||||| | :|: :  | |:|    |||| :| :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 61

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                ||| |     |::|||: :|| |||||||| |    |   |||||||||:||||||:
Db         62 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 118

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:||||| ||||||| |:|:|||||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
                | ::| :  :  | |||||  |  ::||:|||::|: :||:  ||::|  || |
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :       | : :   |       : |:|  :||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
              | ||||:: ::     |:||| | :::::|: ::|:|: |    :   ||:  :: |
Db        298 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 354

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
              |    : : :|  ::|  :  |::     : : :||| ||.|:|    |
Db        355 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 404

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              | || ::::|      |||:|:||| |:|| |
Db        405 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 10
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             33.2%;  Score 842.5;  DB 1;  Length 1296;
  Best Local Similarity   40.6%;  Pred. No. 1.8e-56;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:||||| | :|: :  | |:|    |||| :| :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 919

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                ||| |     |::|||: :|| |||||||| |    |   |||||||||:||||||:
Db        920 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 976

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
```

```
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264

Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321

Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 1155

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380

Db       1156 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 1212

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439

Db       1213 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 1262

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472

Db       1263 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 11
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match           33.2%;  Score 842.5;  DB 2;  Length 1296;
  Best Local Similarity   40.6%;  Pred. No. 1.8e-56;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85

Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 919

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145

Db        920 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 976

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204

Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264

Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321

Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 1155

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380

Db       1156 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 1212

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439

Db       1213 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 1262

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472

Db       1263 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 12
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  PIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             33.2%;  Score 842.5;  DB 3;  Length 1296;
  Best Local Similarity   40.6%;  Pred. No. 1.8e-56;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:||||| | :|: :  | |:|    |||| :| :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 919

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                ||| |     |::|||: :|| |||||||||| |    |   |||||||||:||||||:
Db        920 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN--NEYTIINCMENNSGWKV 976

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:|||| ||||||| |:|:|||||| |
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
              | ::| :  :  | ||||| |  ::||:|||::|: :||:  ||::|  || |
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :      | : :  |      : |:|  :||
Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 1155

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
              | ||||:: ::     |:||| | :::::|: ::|:|: |    :   ||:  :: |
Db       1156 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 1212

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
              |     : : :|  ::|  :  |::     : : :||| ||:|:|     |
Db       1213 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 1262

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              | || ::::|      |||:|:||| |:|| |
Db       1263 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 13
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             33.2%;  Score 842.5;  DB 4;  Length 1296;
  Best Local Similarity   40.6%;  Pred. No. 1.8e-56;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:|||| | :|: :  | |:|    |||| :| :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 919

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                ||| |     |::|||: :|| |||||||||| |    |   |||||||||:||||||:
Db        920 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 976

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:|||| ||||||| |:|:|||||| |
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
              | ::| :  :  | ||||| |  ::||:|||::|: :||:  ||::|  || |
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :      | : :  |      : |:|  :||
Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 1155
```

```
Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
Db       1156 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 1212

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
Db       1213 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 1262

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
Db       1263 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 14
US-07-618-312A-2
; Sequence 2, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
;   APPLICANT:  Makoff Dr, Andrew J
;   APPLICANT:  Romanos Dr, Michael A
;   APPLICANT:  Clare Dr, Jeffrey J
;   APPLICANT:  Fairweather Dr, Neil F
;   TITLE OF INVENTION:  VACCINES
;   NUMBER OF SEQUENCES:  13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  14th Floor
;     STREET:  2200 Clarendon Boulevard,
;     CITY:  Arlington,
;     STATE:  Virginia
;     COUNTRY:  U.S.A.
;     ZIP:  22201
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/618,312A
;     FILING DATE:  19910516
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 8926832.0
;     FILING DATE:  28-NOV-1989
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 9006097.1
;     FILING DATE:  17-MAR-1990
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Crawford Mr, Arthur R
;     REGISTRATION NUMBER:  25,327
;     REFERENCE/DOCKET NUMBER:  510-51
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  0101 703 8750400
;     TELEFAX:  0101 703 5253468
;     TELEX:  200797 NIXN UR
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  452 amino acids
;     TYPE:  AMINO ACID
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             23.9%;  Score 608;  DB 1;  Length 452;
  Best Local Similarity   32.4%;  Pred. No. 5.3e-39;
  Matches  143;  Conservative   87;  Mismatches  183;  Indels   28;  Gaps   12;

Qy         46 ILNLRYRDNNLIDLSGYGAKVEVYDGVKL----NDKNQFKLTSSADSKIRVTQNQNIIFN 101
Db         23 ILNLDINNDINIDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 81

Qy        102 SMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
Db         82 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 141

Qy        157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGE 214
Db        142 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 200

Qy        215 VIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLM 274
Db        201 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 260

Qy        275 YNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNS 334
Db        261 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 320

Qy        335 QSINDDIVRKEDYIHLDLVLHHEEWRVYAY----KYFKEQEEKLFLSIISDSNEFYKTIE 390
Db        321 NEI-DSFVKSGDFIKLYVSYNNNE-HIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 378

Qy        391 IKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKR 450
Db        379 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 433

Qy        451 KPYKSNLGCNWQFIPKDEGWT 471
Db        434 K----ILGCDWYFVPTDEGWT 450


RESULT 15
US-08-280-228-2
; Sequence 2, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
;   APPLICANT:  Makoff Dr, Andrew J
;   APPLICANT:  Romanos Dr, Michael A
;   APPLICANT:  Clare Dr, Jeffrey J
;   APPLICANT:  Fairweather Dr, Neil F
;   TITLE OF INVENTION:  VACCINES
;   NUMBER OF SEQUENCES:  13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  NIXON & VANDERHYE P.C.
;     STREET:  1100 No. 5571694th Glebe Road
;     CITY:  Arlington,
;     STATE:  Virginia
;     COUNTRY:  U.S.A.
;     ZIP:  22201-4714
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/280,228
;     FILING DATE:  25-JUL-1994
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/618,312
;     FILING DATE:  27-NOV-1990
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 8926832.0
;     FILING DATE:  28-NOV-1989
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  GB 9006097.1
;     FILING DATE:  17-MAR-1990
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Wilson, Mary J.
;     REGISTRATION NUMBER:  32,955
;     REFERENCE/DOCKET NUMBER:  117-163
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703) 816-4000
;     TELEFAX:  (703) 816-4100
;     TELEX:  200797 NIXN UR
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-280-228-2

  Query Match             23.9%;  Score 608;  DB 1;  Length 452;
  Best Local Similarity   32.4%;  Pred. No. 5.3e-39;
  Matches  143;  Conservative   87;  Mismatches  183;  Indels   28;  Gaps   12;

Qy         46 ILNLRYRDNNLIDLSGYGAKVEVYDGVKL----NDKNQFKLTSSADSKIRVTQNQNIIFN 101
              ||||   ::  ||:||: : |  |   :|    | |    | ::  |:: | :  :| :|
Db         23 ILNLDINNDINIDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 81

Qy        102 SMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
               || :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :||||
Db         82 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 141

Qy        157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGE 214
              |  |: : : |  :: :  : |: |:| |:||||: | :| :|||| |  : :|  :|
Db        142 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 200

Qy        215 VIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLM 274
              :  :  || |||   :  |:: :  | ||   ||   |:::|       :|:||||||
Db        201 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 260

Qy        275 YNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNS 334
              |: |||:    : :  ::|   :    :     |       | || || | ||||:| : :
Db        261 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 320

Qy        335 QSINDDIVRKEDYIHLDLVLHHEEWRVYAY----KYFKEQEEKLFLSIISDSNEFYKTIE 390
              | |  |:  |:| | :  :: |  :  |      |   :  | :   :     || :|
Db        321 NEI-DSFVKSGDFIKLYVSYNNNE-HIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 378

Qy        391 IKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKR 450
              :  :  :|| ||    |: ::  :||:| |      :     :|    | ||   :|
Db        379 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 433

Qy        451 KPYKSNLGCNWQFIPKDEGWT 471
              |     |||:| |:| |||||
Db        434 K----ILGCDWYFVPTDEGWT 450


Search completed: September 10, 2004, 03:46:54
Job time : 12.0183 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 91.3645 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 2540
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2540	100.0	472	12	US-10-272-898-44	Sequence 44, Appl
2	2540	100.0	472	15	US-10-354-774-44	Sequence 44, Appl
3	2540	100.0	472	15	US-10-271-012-44	Sequence 44, Appl
4	2540	100.0	472	16	US-10-729-122-44	Sequence 44, Appl
5	2540	100.0	472	16	US-10-729-039-44	Sequence 44, Appl
6	2358	92.8	1291	12	US-10-272-898-40	Sequence 40, Appl
7	2358	92.8	1291	15	US-10-354-774-40	Sequence 40, Appl
8	2358	92.8	1291	15	US-10-271-012-40	Sequence 40, Appl
9	2358	92.8	1291	15	US-10-452-024-117	Sequence 117, App
10	2358	92.8	1291	15	US-10-205-516-4	Sequence 4, Appli
11	2358	92.8	1291	16	US-10-729-122-40	Sequence 40, Appl
12	2358	92.8	1291	16	US-10-729-039-40	Sequence 40, Appl
13	2358	92.8	1301	15	US-10-205-516-18	Sequence 18, Appl
14	2307	90.8	472	12	US-10-272-898-46	Sequence 46, Appl
15	2307	90.8	472	15	US-10-354-774-46	Sequence 46, Appl
16	2307	90.8	472	15	US-10-271-012-46	Sequence 46, Appl
17	2307	90.8	472	16	US-10-729-122-46	Sequence 46, Appl
18	2307	90.8	472	16	US-10-729-039-46	Sequence 46, Appl
19	2145	84.4	1291	15	US-10-452-024-142	Sequence 142, App
20	2144	84.4	1290	15	US-10-360-101-220	Sequence 220, App
21	2144	84.4	1290	15	US-10-452-024-119	Sequence 119, App
22	2144	84.4	1291	12	US-10-272-898-42	Sequence 42, Appl
23	2144	84.4	1291	15	US-10-354-774-42	Sequence 42, Appl
24	2144	84.4	1291	15	US-10-271-012-42	Sequence 42, Appl
25	2144	84.4	1291	15	US-10-452-024-2	Sequence 2, Appli
26	2144	84.4	1291	15	US-10-452-024-118	Sequence 118, App
27	2144	84.4	1291	16	US-10-729-122-42	Sequence 42, Appl
28	2144	84.4	1291	16	US-10-729-039-42	Sequence 42, Appl
29	2119	83.4	1291	15	US-10-452-024-121	Sequence 121, App
30	2113	83.2	1291	15	US-10-452-024-123	Sequence 123, App
31	2108	83.0	1291	15	US-10-452-024-122	Sequence 122, App
32	2102	82.8	440	10	US-09-910-186A-8	Sequence 8, Appli
33	2090	82.3	441	15	US-10-452-024-188	Sequence 188, App
34	1586	62.4	1169	14	US-10-241-596-20	Sequence 20, Appl
35	1350.5	53.2	473	12	US-10-272-898-79	Sequence 79, Appl
36	1350.5	53.2	473	15	US-10-354-774-79	Sequence 79, Appl
37	1350.5	53.2	473	15	US-10-271-012-79	Sequence 79, Appl
38	1350.5	53.2	473	16	US-10-729-122-79	Sequence 79, Appl
39	1350.5	53.2	473	16	US-10-729-039-79	Sequence 79, Appl
40	1191.5	46.9	449	10	US-09-910-186A-18	Sequence 18, Appl
41	1187.5	46.8	1297	12	US-10-272-898-77	Sequence 77, Appl
42	1187.5	46.8	1297	15	US-10-354-774-77	Sequence 77, Appl
43	1187.5	46.8	1297	15	US-10-271-012-77	Sequence 77, Appl
44	1187.5	46.8	1297	15	US-10-452-024-7	Sequence 7, Appli
45	1187.5	46.8	1297	15	US-10-452-024-150	Sequence 150, App

ALIGNMENTS

```
RESULT 1
US-10-272-898-44
; Sequence 44, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;        SEQUENCE CHARACTERISTICS:
```

```
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-272-898-44

  Query Match            100.0%;  Score 2540;  DB 12;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-187;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 2
US-10-354-774-44
; Sequence 44, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-354-774-44

  Query Match            100.0%;  Score 2540;  DB 15;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-187;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 3
US-10-271-012-44
; Sequence 44, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-271-012-44

  Query Match             100.0%;  Score 2540;  DB 15;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-187;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 4
US-10-729-122-44
; Sequence 44, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-729-122-44

  Query Match             100.0%;  Score 2540;  DB 16;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-187;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 5
US-10-729-039-44
```

```
; Sequence 44, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 44:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-729-039-44

  Query Match             100.0%;  Score 2540;  DB 16;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-187;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472


RESULT 6
US-10-272-898-40
; Sequence 40, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1291 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-272-898-40

  Query Match             92.8%;  Score 2358;  DB 12;  Length 1291;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-172;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
                ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 7
US-10-354-774-40
; Sequence 40, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-354-774-40

  Query Match             92.8%;  Score 2358;  DB 15;  Length 1291;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-172;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
                |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
                ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 8
US-10-271-012-40
; Sequence 40, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-271-012-40

  Query Match             92.8%;  Score 2358;  DB 15;  Length 1291;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-172;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291
```

RESULT 9
US-10-452-024-117
; Sequence 117, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 117
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-117

Query Match 92.8%; Score 2358; DB 15; Length 1291;
Best Local Similarity 99.5%; Pred. No. 2.7e-172;
Matches 442; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291
```

RESULT 10
US-10-205-516-4
; Sequence 4, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-4

Query Match 92.8%; Score 2358; DB 15; Length 1291;
Best Local Similarity 99.5%; Pred. No. 2.7e-172;
Matches 442; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291
```

```
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 11
US-10-729-122-40
; Sequence 40, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-729-122-40

  Query Match            92.8%;  Score 2358;  DB 16;  Length 1291;
  Best Local Similarity  99.5%;  Pred. No. 2.7e-172;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 12
US-10-729-039-40
; Sequence 40, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-729-039-40

  Query Match            92.8%;  Score 2358;  DB 16;  Length 1291;
  Best Local Similarity  99.5%;  Pred. No. 2.7e-172;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027
```

```
Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 13
US-10-205-516-18
; Sequence 18, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 1301
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-18

  Query Match             92.8%;  Score 2358;  DB 15;  Length 1301;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-172;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 911

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 971

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1031

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1091

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1151

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1211

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1271

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1272 KRKPYKSNLGCNWQFIPKDEGWTE 1295


RESULT 14
US-10-272-898-46
; Sequence 46, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-272-898-46

  Query Match             90.8%;  Score 2307;  DB 12;  Length 472;
  Best Local Similarity   89.8%;  Pred. No. 6.1e-169;
  Matches  424;  Conservative   26;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||:|||||||||||||:|||||||||||||||:|||||||||||||||:|
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||:|||||||| |||| ||||| ||| |||| |||||||:||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              |||||:|:||||:| ||||||||||||||||||||||||||||||||| ||| ||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
```

```
Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||| |||||||| |||||:||||||||||:||||||||||||||||:||    ::|||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              || |||||::||||||: ||||:||| ||:||||||||:||||||||||||||:|||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||:: ::|||||||||||||||||    ||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
```

```
RESULT 15
US-10-354-774-46
; Sequence 46, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-354-774-46

  Query Match             90.8%;  Score 2307;  DB 15;  Length 472;
  Best Local Similarity   89.8%;  Pred. No. 6.1e-169;
  Matches  424;  Conservative   26;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120
              ||||||||||:|||||||||||:|||||||||||||:|||||||||||||:|
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240
              ||||||||||||:|||||||| |||| ||||| ||| |||| |||||||:||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300
              |||||:|:||||:| ||||||||||||||||||||||||||||||||||| ||| ||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
              ||| |||||||| |||||:||||||||||:||||||||||||||||:||    ::|||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420
              || |||||::||||||: ||||:||| ||:||||||||:||||||||||||||:|||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||:: ::|||||||||||||||||    ||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
```

Search completed: September 10, 2004, 04:57:59
Job time : 92.3645 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 8.33859 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 2540
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2358	92.8	1291	2	I40631	non-proteolytic bo
2	2144	84.4	1291	1	A48940	bontoxilysin (EC 3
3	1187.5	46.8	1297	2	S39791	neurotoxin - Clost
4	842.5	33.2	1296	1	BTCLAB	bontoxilysin (EC 3
5	798	31.4	1268	2	S33411	botulinum neurotox
6	790.5	31.1	1296	2	I40645	botulinum neurotox
7	752	29.6	1274	2	I40813	neurotoxin type F
8	725.5	28.6	1252	2	S21178	botulinum neurotox
9	708	27.9	1251	2	JH0256	botulinum neurotox
10	614.5	24.2	1276	2	S11455	botulinum neurotox
11	607	23.9	1315	1	BTCLTN	tentoxilysin (EC 3
12	565	22.2	1291	2	S46431	botulinum neurotox
13	565	22.2	1291	2	A49777	botulinum neurotox
14	563	22.2	1285	2	S70582	botulinum neurotox
15	351	13.8	366	2	S48110	neurotoxin type F
16	328	12.9	369	2	S48109	neurotoxin type F
17	324.5	12.8	367	2	S48106	neurotoxin type E
18	241	9.5	1162	2	I40817	botulinum toxin no
19	240.5	9.5	1162	2	A47708	progenitor toxin n
20	215.5	8.5	1196	2	JQ1467	toxin, nontoxic co
21	215.5	8.5	1196	2	S46430	botulinum neurotox
22	207.5	8.2	1165	2	I40644	botulinum neurotox
23	203.5	8.0	1193	2	S68218	botulinum neurotox
24	191.5	7.5	1193	2	JC4901	nontoxic-nonhemagg
25	184.5	7.3	2485	1	H71621	serine/threonine-s
26	170.5	6.7	1061	2	D98008	conserved hypothet
27	170	6.7	1553	2	T18502	hypothetical prote
28	166.5	6.6	960	2	S72284	DNA-directed RNA p
29	163.5	6.4	789	2	G90587	lipoprotein [impor
30	162.5	6.4	763	2	S17998	gene COX1 intron 4
31	161	6.3	1121	2	F71613	hypothetical prote
32	160	6.3	1494	2	C70152	sensory transducti
33	160	6.3	2013	2	C71610	probable membrane
34	159	6.3	3119	2	T18414	protein g377 - mal
35	158.5	6.2	1855	2	S41649	DNA polymerase - m
36	158.5	6.2	3394	2	T18501	hypothetical prote
37	158.5	6.2	3973	2	B71612	hypothetical prote
38	158	6.2	336	2	S04682	ribosomal protein
39	156.5	6.2	1802	2	G71616	hypothetical prote
40	156	6.1	1127	2	T28317	ORF MSV156 hypothe
41	156	6.1	4981	2	T18489	hypothetical prote
42	155.5	6.1	787	2	H90543	conserved hypothet
43	155.5	6.1	2539	2	B71619	hypothetical prote
44	155	6.1	1516	2	E71619	RAD2 endonuclease
45	154.5	6.1	1182	2	G71607	probable integral

ALIGNMENTS

RESULT 1
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994.
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinur
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 92.8%; Score 2358; DB 2; Length 1291;
Best Local Similarity 99.5%; Pred. No. 3.3e-125;
Matches 442; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291
```

RESULT 2
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisio
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match   84.4%; Score 2144; DB 1; Length 1291;
  Best Local Similarity   89.2%; Pred. No. 3.8e-113;
  Matches 398; Conservative   26; Mismatches   22; Indels   0; Gaps   0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              |||||||||||||||||||||||:|||||||||||||||||||:||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|||||||||||||:|||||||||||||:|| ||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||||| ||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
               ||||| ||| |||| ||||||:|||||||||||||||:|:||||:| ||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              ||||||||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    ::||||| |||||::||||||: ||||:|||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              ||:||||||||:|||||||||||||||:||||||||||||:: ::||||||||||||||
```

```
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||| ||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 3
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             46.8%;  Score 1187.5;  DB 2;  Length 1297;
  Best Local Similarity   50.2%;  Pred. No. 2.9e-59;
  Matches  227;  Conservative   94;  Mismatches  120;  Indels   11;  Gaps    7;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
              |||||::|| | | | :| ||:| ||   ||| ||||| : |   | ||    |||| :
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWK 144
              | :| |   |:: ::::||| :||::||:| ||| |:||| |: |||||:|:||:|||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTL 203
              :||:|||||||||:| |:||:||||:|:::||:|||:|| :||||: | || |||||:|
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        204 ESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSE 263
              : :  | ::  :  : :| |||    | |:|:|:| |:||  :|| : :  :| ||| :
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIG 323
              ||||||||| |: :||:|| | :| |||    :|:||   |: :|  :  |||:|||:|
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 1148

Qy        324 EKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQEEKLFLSIISD 381
              :|||:: |||::| ||:|||: |||:|::  :  | :|||     || : :|||: |:|
Db       1149 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 1208

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GVLRKKYKDYFCI 440
                ||  ::||:| |: :|:|::| :||   |   || || :| :  | :   | :||||
Db       1209 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 1265

Qy        441 SKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |:|||: :     |  |||||||| ||||||
Db       1266 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 4
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin,
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ext
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arour
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; N
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prote
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

  Query Match             33.2%;  Score 842.5;  DB 1;  Length 1296;
```

```
  Best Local Similarity   40.6%;  Pred. No. 7.7e-40;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:|||||  | :|: :  | |:|    |||| :| :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 919

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
               ||| |     |::|||: :|| ||||||||| |     |   ||||||||||:||||||:
Db        920 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 976

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |      : | |:|:   :||:|||||  |||||| |:|:||||||  |
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264
              | ::| :  :  | |||||  |  ::||:|||::|: :||:  ||::|  || |
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321
              ||||||: | |:| ||| |  : | |: :       | : :   |       : |:|  :||
Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 1155

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380
              | ||||:: ::     |:|||  | :::::|: ::|:|: |    :   ||:  :: |
Db       1156 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 1212

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439
              |     : : :|  ::|  :  |::      : : :||| || |:|      |
Db       1213 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 1262

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              | || ::::|      |||:|:||| |:|| |
Db       1263 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 5
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match           31.4%;  Score 798;  DB 2;  Length 1268;
  Best Local Similarity   38.6%;  Pred. No. 2.4e-37;
  Matches  174;  Conservative   95;  Mismatches  152;  Indels   30;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | ||| :  ::   |::: |||::| :|  || ||||: : :   : :   ::||| : |
Db        843 DKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYS 902

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: :|||  ||:|| : :||||||:||||| |  ::| ::|||||||||: |||||
Db        903 SRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN--LKN-LNNEYTIINCMRNNNSGW 959

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |||:  | ||||| |  |  : : | |    |||:|||:| ||||||| | ::|:||||
Db        960 KISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGN 1019

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |     | ::| : |:  | ||: |  | |::: ::|| ||| :|::: |: :|  :  |
Db       1020 LTDQKSILNLGNIHVDDNILFKIVGCND-TRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              ||||||| |:|||:||: |    |  : : |:| :  |  :       :|  :   ||
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPN--MSVTKNSDILNINRQRGIYSKTNIFSNARLYT 1136

Qy        323 GEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISD 381
              | : |||:  :: : | |: ||| | :::::|  : |:::||       |: : |  ||:
Db       1137 GVEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRISN 1196

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              ||  | : ::    |     :| : || :  : :||||:| |             :   |
Db       1197 SN--YNSNQMIIMDSIGD-NCTMNFKTN--NGNDIGLLGFH-----------LNNLVAS 1239

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              || | ::      | || | || |: || |
Db       1240 SWYYKNIRNN--TRNNGCFWSFISKEHGWQE 1268


RESULT 6
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A a
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match           31.1%;  Score 790.5;  DB 2;  Length 1296;
  Best Local Similarity   38.8%;  Pred. No. 6.6e-37;
  Matches  175;  Conservative   89;  Mismatches  160;  Indels   27;  Gaps   12;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN--DKNQFKLTSSAD 87
              |:  | :|   |:|  ||:: |: ::||||| ||||: : | |  :  |||| || :
Db        862 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 921

Qy         88 SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISI 147
              | | |     |:|||: :|| ||||:|||| :   :  ::|||||||::|||||:|:
Db        922 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLESN 206
              |||||  |     : | |:|:   :||:||||| ||||||  |  :||||||  |
Db        979 NYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQ 1038

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              | ::| :  : :| |||||  |  ::| :|||::|: :||:  ||::|  || |  ||
Db       1039 KPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLYIG 323
              ||||| | |:| ||| |  : | |: :       | : :   |       : |:|   || |
Db       1099 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLN-STLYEG 1157

Qy        324 EKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-ISDS 382
              ||||:: ::     |:|||  | :::::|: ::|:|: |    :   ||:  :: | |
Db       1158 TKFIIKKYASGN--EDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIPDV 1214

Qy        383 NEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              : : :|  |:|     |::     : : :||| || | :   |                 |
Db       1215 GNLSQVVVMKSKDDQGIRNKCKM--NLQDNNGNDIGFIGFHLYDNIAKL--------VAS 1264

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              || ::| :       ||:|:||| |:|| |
Db       1265 NWYNRQVGKA--SRTFGCSWEFIPVDDGWGE 1293
```

RESULT 7
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 29.6%; Score 752; DB 2; Length 1274;
Best Local Similarity 37.7%; Pred. No. 9.5e-35;
Matches 171; Conservative 93; Mismatches 152; Indels 38; Gaps 15;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  ||:   :| :: ||::|| :|  ||:|||: : :   | :   ::||| : :
Db        851 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 910

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCM-KNNSGW 143
              |  |:: : || :||:|| : :||:|||:||||:       |    |||||||  |||||
Db        911 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGW 967

Qy        144 KISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYI 199
              |||:|  |   |||||  | :|  ::: | |     || |||:| |||||||  | |::|||
Db        968 KISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYI 1027

Qy        200 NGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQ 259
              || |     | ::|:: |:  | ||: |  | | :: ::|| :|||:|::: |: :|  :
Db       1028 NGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNE 1086

Qy        260 SYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRN 319
                ||::||| |:|||:||:||   |: || |  :| :  |  :    : | ::||:
Db       1087 PDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK- 1143

Qy        320 LYIGEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI 378
              || | : |||:       | |: ||| |  ::::|    |:|:||    | ::||:  :
Db       1144 LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIRT- 1200

Qy        379 ISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYF 438
              |: |:    | :    :    :| : |: :  |  :|||:| |              :
Db       1201 -SNLNDSLGQIIVM---DSIGNNCTMNFQNNNGS--NIGLLGFH-----------SNNL 1242

Qy        439 CISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
                | ||   ::|    |: || |  | |: || |
Db       1243 VASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 1274
```

RESULT 8
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

Query Match 28.6%; Score 725.5; DB 2; Length 1252;
Best Local Similarity 35.5%; Pred. No. 2.9e-33;
Matches 160; Conservative 99; Mismatches 155; Indels 37; Gaps 13;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:||:::  :| ||| : : :   |     :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
                |:: ::||  ||::: : :||:|||:||| | |  :   ::||||||||: |||||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |:|:  | ||||| |  |  : : | |      ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |    | ::| : |:  | ||:  :   |::| ::||:||: :|::: |: :|  :  :
Db       1011 LIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              ||||||| |:|:||||: |    |::|   |||::      |  |  |  :    ||
```

```
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTL------SINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEE-WRVYAYKYFKEQEEKLFLSIISD 381
              | |  |:| :|| | ||::||| | :::: |       : :||      :|: : :|  |
Db       1124 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKIS--SS 1180

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              | | : : :          :| : || :  : ::|||:|                |    |
Db       1181 GNRFNQVVVMNSVGN----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1222

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
               ||   :  : : :: || | || :: || |
Db       1223 TWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251


RESULT 9
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match            27.9%;  Score 708;  DB 2;  Length 1251;
  Best Local Similarity  35.7%;  Pred. No. 2.8e-32;
  Matches  161;  Conservative   95;  Mismatches  157;  Indels   38;  Gaps   14;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:||:::  :| ||| : : :    |     :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
                |:: ::||  ||::: : :||:|||:||| | |  :   ::||||||||: |||||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |:|:  | ||||| | :|  : : | |      ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::| : |:  | ||:  :   |::| ::||:||: :|::: |: :|  : :
Db       1011 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              |||||||  |:|:|||| |    |::|    ||::      |  |  |  :    ||
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTL------SINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEE-WRVYAYKYFKEQEEKLFLSIISD 381
              | |  |:| :|| | ||::||| | :::: |       :||      :|: : :|  |
Db       1124 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKIS--SS 1180

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              | | : :          :| : || :  : ::|||:|                |    |
Db       1181 GNRFNQVVVMNSVG-----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1221

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
               ||   : |    || |  | || :: || |
Db       1222 TWYYTHM-RDNTNSN-GFFWNFISEEHGWQE 1250


RESULT 10
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Krieglst
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type I
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            24.2%;  Score 614.5;  DB 2;  Length 1276;
  Best Local Similarity  32.9%;  Pred. No. 5.3e-27;
  Matches  157;  Conservative   89;  Mismatches  150;  Indels   81;  Gaps   16;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
              :::| :: |:| : | :: ||:|: : | |:| ||| |:| | | |:||    | |||:|
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWK 144
              | | || |  | ||:::::: : ||||||:| |    |: |  |||||||| :: |||||
Db        910 SGD-KIIVNLNNNILYSAIYENSSVSFWIKISK----DLTN-SHNEYTIINSIEQNSGWK 963

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTL 203
              : ||   | | | |:| | ||: |:|:     : | |:|||||||| :   |:|||| |
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        204 ESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSE 263
              : :  |:|: || ::  | | :| ::|  | :|:: |:||: :|:  :|  :|: |
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIG 323
              :||:||||| :: |||: |    : ||     :    :|:  :|   |       || |
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVQYPDRS------KLYTG 1132

Qy        324 EKFIIRRESN----SQSINDDIVRKEDYIHLDLVLHHEEWRVY-AYKYFKEQEEKLFLSI 378
                 |:  |:    |: :| |           :::||      :| : ||           |
Db       1133 NPITIKSVSDKNPYSRILNGD----------NIILH----MLYNSRKYM----------I 1168

Qy        379 ISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKK----- 433
              | |::  | | :   | :   |: :|              ::|  ||  |    :: |
Db       1169 IRDTDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQ 1219

Qy        434 ---------------YKDYFCISKWYLKEVKRKPYKSNL---GCNWQFIPKDEGWTE 472
                            || :    |    |    |:: |      |:|| :| || |
Db       1220 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 11
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
```

```
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match           23.9%;  Score 607;  DB 1;  Length 1315;
  Best Local Similarity  32.2%;  Pred. No. 1.5e-26;
  Matches  142;  Conservative   88;  Mismatches  183;  Indels   28;  Gaps   12;

Qy        46 ILNLRYRDNNLIDLSGYGAKVEVYDGVKL----NDKNQFKLTSSADSKIRVTQNQNIIFN 101
             ||||   :: : |:||: : |  |   :|    | |    | ::  |:: | :  :| :|
Db       886 ILNLDINNDIISDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 944

Qy       102 SMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
             || :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :||||
Db       945 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 1004

Qy       157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGE 214
             |  |: : : |  :: :  : |: |:| |:|||: | :| :|||| |  : :|  :|
Db      1005 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 1063

Qy       215 VIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLM 274
             :  :  || |||   :  |:: :  | ||   ||   |:::|       :|:|||||||
Db      1064 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 1123

Qy       275 YNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNS 334
             |: |||:    : :  ::|  :   :      |       | || || | ||||:| : :
Db      1124 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 1183

Qy       335 QSINDDIVRKEDYIHLDLVLHHEEWRVYAY----KYFKEQEEKLFLSIISDSNEFYKTIE 390
               |  |:  |:| | :  :: |  :  |      |  :  | :   :     || :|
Db      1184 NEI-DSFVKSGDFIKLYVSYNNNE-HIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 1241

Qy       391 IKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKR 450
               :  :  :|| ||    |: ::  :||:| |      :     :|    | ||   :|  .
Db      1242 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 1296

Qy       451 KPYKSNLGCNWQFIPKDEGWT 471
             |     |||:| |:| |||||
Db      1297 K----ILGCDWYFVPTDEGWT 1313


RESULT 12
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

  Query Match           22.2%;  Score 565;  DB 2;  Length 1291;
  Best Local Similarity  32.4%;  Pred. No. 3.3e-24;
  Matches  156;  Conservative   87;  Mismatches  155;  Indels   84;  Gaps   20;

Qy        27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
             :::| :: |:| : | :: ||:|: | | |:| ||| |:|     |:||      ||| |
Db       854 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 913

Qy        85 SAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 142
             | :   |: ||||:||::|||:  ||:|||||| |:    : |     ||||: :|||||
Db       914 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSG 967

Qy       143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 201
             | | |  | :::||       :|: | |:|  :   | |:|||||:||| : | ||||||
Db       968 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1026

Qy       202 TLESNMDIKDIGEVIVNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQSNI 252
```

```
Db       1027 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDI 1085

Qy        253 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNS 312

Db       1086 NILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN- 1140

Qy        313 NYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEE 372

Db       1141 ------DFNEGYKIIIKRIRG--NTNDTRVRGGDILYFDMTINNK-----AYNLFMKNE- 1186

Qy        373 KLFLSIISD--SNEFYKTIEIKE----------YDEQPS----YSCQLLFKK--DEESTD 414

Db       1187 ----TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 1242

Qy        415 DIGLIGIHRFYESGVLRKKYKDYFCISKWY-----LKEVKRKPYKSNL---GCNWQFIPK 466

Db       1243 GICSIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPV 1289

Qy        467 DE 468

Db       1290 SE 1291
```

RESULT 13
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 22.2%; Score 565; DB 2; Length 1291;
  Best Local Similarity 32.4%; Pred. No. 3.3e-24;
  Matches 156; Conservative 87; Mismatches 155; Indels 84; Gaps 20;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84

Db        854 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 913

Qy         85 SAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 142

Db        914 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSG 967

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 201

Db        968 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1026

Qy        202 TLESNMDIKDIGEVIVNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQSNI 252

Db       1027 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDI 1085

Qy        253 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNS 312

Db       1086 NILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN- 1140

Qy        313 NYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEE 372

Db       1141 ------DFNEGYKIIIKRIRG--NTNDTRVRGGDILYFDMTINNK-----AYNLFMKNE- 1186

Qy        373 KLFLSIISD--SNEFYKTIEIKE----------YDEQPS----YSCQLLFKK--DEESTD 414

Db       1187 ----TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 1242

Qy        415 DIGLIGIHRFYESGVLRKKYKDYFCISKWY-----LKEVKRKPYKSNL---GCNWQFIPK 466

Db       1243 GICSIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPV 1289

Qy        467 DE 468

Db       1290 SE 1291
```

RESULT 14
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C and
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membra
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match 22.2%; Score 563; DB 2; Length 1285;
  Best Local Similarity 32.2%; Pred. No. 4.2e-24;
  Matches 151; Conservative 92; Mismatches 166; Indels 60; Gaps 19;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84

Db        850 NSLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGS 909

Qy         85 SAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 142

Db        910 SGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSG 963

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 201
```

```
Db        964 WSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGY-NKWFFVTITTNMMGNMMIYING 1022

Qy        202 TLESNMDIKDIGEVIVNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQSNI 252
              |    : :|::  :  :  |||:::          | |    :|:: | ||  :|:  :|
Db       1023 KLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNIN-MWIRDFYIFAKELDDKDI 1081

Qy        253 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNS 312
              ::    |:  :||:||| | |:|||| |    | |:    |  |  ::   :|
Db       1082 NILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMNRYM-----SKKGNGIV---FNTRK 1133

Qy        313 NYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEE 372
              |  ::    | | ||:|     : ||  || |: :: :  : :::: :  ||  :
Db       1134 NNNDFNE---GYKIIIKRIRG--NTNDTRVRGENVLYFNTTIDNKQYSLGMYKPSRNLGT 1188

Qy        373 KLF-LSIISDSNEFYKTIEIKEYDE---QP----SYSCQLLFKKDEESTDDIGL--IGIH 422
              |  |  :   :    ||::|     ||     |    ||     :|: :|:  || :
Db       1189 DLVPLGALDQPMD-----EIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSY 1243

Qy        423 RFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNL---GCNWQFIPKDE 468
              |       |   ||:   :: :  :| : | | |     :| |:|  |
Db       1244 SF-------KLGDDYWFNHEYLIPVIKIEHYASLLESTSTHWVFVPASE 1285


RESULT 15
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             13.8%;  Score 351;  DB 2;  Length 366;
  Best Local Similarity   46.3%;  Pred. No. 7.6e-13;
  Matches   68;  Conservative   32;  Mismatches   41;  Indels    6;  Gaps    3;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||    :| :| ||::|| :|  ||:||||: : :   | :   ::||| : |
Db        218 DKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYS 277

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: : || :||:|  : :||:|||:||||| |   :  ::|||||:|:: |||||
Db        278 SKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFN---KVNLNNEYTIIDCIRNNNSGW 334

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEY 170
              |||:  |:||||| |  |  : : | |
Db        335 KISLNYNKIIWTLQDTAGNNQKLVFNY 361


Search completed: September 10, 2004, 03:43:55
Job time : 10.3386 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 5.75902 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 2540
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2144	84.4	1290	1	BXB_CLOBO	P10844 clostridium
2	1187.5	46.8	1296	1	BXG_CLOBO	Q60393 clostridium
3	842.5	33.2	1295	1	BXA1_CLOBO	P10845 clostridium
4	790.5	31.1	1295	1	BXA2_CLOBO	Q45894 clostridium
5	752	29.6	1274	1	BXF_CLOBO	P30996 clostridium
6	723	28.5	1250	1	BXE_CLOBO	Q00496 clostridium
7	708	27.9	1250	1	BXE_CLOBU	P30995 clostridium
8	614.5	24.2	1276	1	BXD_CLOBO	P19321 clostridium
9	607	23.9	1314	1	TETX_CLOTE	P04958 clostridium
10	565	22.2	1290	1	BXC1_CLOBO	P18640 clostridium
11	241	9.5	1162	1	BXEN_CLOBU	Q06366 clostridium
12	240.5	9.5	1162	1	BXEN_CLOBO	P46082 clostridium
13	215.5	8.5	1196	1	BXCN_CLOBO	P46081 clostridium
14	158	6.2	339	1	RMAR_CANGA	P21358 candida gla
15	150	5.9	360	1	MTLD_MYCPU	Q98ph2 mycoplasma
16	149.5	5.9	2136	1	YCF2_MARPO	P09975 marchantia
17	148	5.8	964	1	DPOL_CBEPV	P30319 choristoneu
18	146.5	5.8	445	1	II2A_MOUSE	P15091 mus musculu
19	146.5	5.8	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
20	145	5.7	817	1	RPOD_ASTLO	P58132 astasia lon
21	144.5	5.7	445	1	II2B_MOUSE	Q9r002 mus musculu
22	142	5.6	611	1	SNF1_CANGA	Q00372 candida gla
23	139.5	5.5	765	1	YG26_METJA	Q59019 methanococc
24	137.5	5.4	544	1	FLGK_BUCAP	Q8k9k0 buchnera ap
25	137.5	5.4	569	1	CYSP_PLAFA	P25805 plasmodium
26	137.5	5.4	626	1	PARC_BORBU	O51066 borrelia bu
27	137.5	5.4	2339	1	RPC1_PLAFA	P27625 plasmodium
28	136.5	5.4	1018	1	YC14_METJA	Q58611 methanococc
29	136	5.4	634	1	YCX3_EUGGR	P31916 euglena gra
30	136	5.4	1024	1	RPOB_PLAFA	P21421 plasmodium
31	135.5	5.3	519	1	YC82_ASTLO	P58145 astasia lon
32	135	5.3	908	1	DPO1_BORBU	O51498 borrelia bu
33	135	5.3	2366	1	TOXB_CLODI	P18177 clostridium
34	135	5.3	2710	1	TOXA_CLODI	P16154 clostridium
35	132	5.2	968	1	Y080_BUCAP	Q8ka32 buchnera ap
36	131	5.2	476	1	RF3_SACBA	P05512 saccharomyc
37	131	5.2	504	1	Y795_METJA	Q58205 methanococc
38	131	5.2	995	1	DPO1_KLULA	P09804 kluyveromyc
39	129.5	5.1	380	1	IPYR_PLAF7	O77392 plasmodium
40	129	5.1	735	1	CIGB_DICDI	Q94481 dictyosteli
41	129	5.1	1288	1	SMC4_HUMAN	Q9ntj3 homo sapien
42	128.5	5.1	580	1	P69_MYCHR	P15362 mycoplasma
43	128	5.0	386	1	RMAR_HANWI	P48849 hansenula w
44	128	5.0	970	1	Y087_BUCAI	P57189 buchnera ap
45	127	5.0	647	1	NTP1_MSEPV	Q9yw39 melanoplus

ALIGNMENTS

RESULT 1
BXB_CLOBO
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence.";
RL Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.";
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;

```
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             84.4%;  Score 2144;  DB 1;  Length 1290;
  Best Local Similarity   89.2%;  Pred. No. 7.8e-113;
  Matches  398;  Conservative   26;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              ||||||||||||||||||||||||:||||||||||||||||||:||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|||||||||||||:|||||||||||||||:|| ||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||| ||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              ||||| ||| |||| |||||:|||||||||||||||:|:||||:| ||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    ::||||| |||||::||||||: ||||:|||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              ||:|||||||:|||||||||||||:||||||||||||:: ::||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||||||   ||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 2
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
```

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             46.8%;  Score 1187.5;  DB 1;  Length 1296;
  Best Local Similarity   50.2%;  Pred. No. 2.7e-59;
  Matches  227;  Conservative   94;  Mismatches  120;  Indels   11;  Gaps    7;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
              |||||::|| | | | :| ||:| ||   ||| ||||| : |   |  ||    |||| :
Db        850 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 909

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWK 144
              | :| |   |:: ::::||| :||::||:| ||| |:||| |: |||||:|:||:||||
Db        910 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 969

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTL 203
             :||:|||||||||:| |:||:||||:|:::||:|||:|| :||||: | || |||||:|
Db        970 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1029

Qy        204 ESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSE 263
              : :  | ::  :  : :| |||    | |:|:|:| |:||  :|| : :  :| ||| :
Db       1030 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1089

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIG 323
              |||||||||  |: :||:|| | :| |||    :|:||   |: :|  :  |||:|||:|
Db       1090 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 1147

Qy        324 EKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQEEKLFLSIISD 381
             :|||:: |||::| ||:|||: |||:|::  :  | :|||     || : :|||: |:|
Db       1148 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 1207

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GVLRKKYKDYFCI 440
                ||  ::||:| |: :|:||:| :||   |   || || :| :  | :   | :||||
Db       1208 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 1264

Qy        441 SKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |:|||: :     |  ||||||||| ||||||
Db       1265 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1296
```

```
RESULT 3
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
```

```
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             33.2%;  Score 842.5;  DB 1;  Length 1295;
  Best Local Similarity   40.6%;  Pred. No. 5.5e-40;
  Matches  184;  Conservative   86;  Mismatches  152;  Indels   31;  Gaps   13;

Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN----DKNQFKLTSS 85
              |:  | :|   |:|  ||||||  |:||||| | :|: :  | |:|    |||| :| :
Db        861 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI--GSKVNFDPIDKNQIQLFNL 918

Qy         86 ADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKI 145
                ||| |     |::|||: :|| |||||||| |    |   |||||||||:||||||:
Db        919 ESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLN---NEYTIINCMENNSGWKV 975

Qy        146 SIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLE 204
              |:    ||||| |     : | |:|:   :||:|||| ||||||| |:|:|||||| |
```

```
Db        976 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1035

Qy        205 SNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEY 264

Db       1036 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1095

Qy        265 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLY 321

Db       1096 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 1154

Qy        322 IGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-IS 380

Db       1155 RGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIP 1211

Qy        381 DSNEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFC 439

Db       1212 DVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIAKL--------V 1261

Qy        440 ISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472

Db       1262 ASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1292
```

RESULT 4
BXA2_CLOBO
ID BXA2_CLOBO STANDARD; PRT; 1295 AA.
AC Q45894; P77780;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kyoto-F;
RX MEDLINE=94143603; PubMed=8310180;
RA Willems A., East A.K., Lawson P.A., Collins M.D.;
RT "Sequence of the gene coding for the neurotoxin of Clostridium
RT botulinum type A associated with infant botulism: comparison with
RT other clostridial neurotoxins.";
RL Res. Microbiol. 144:547-556(1993).
RN [2]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=Kyoto-F;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC binds with high affinity to peripheral neuronal presynaptic
CC membrane, is then internalized by receptor-mediated endocytosis.
CC The C-terminus of the heavy chain (H) is responsible for the
CC adherence of the toxin to the cell surface while the N-terminus
CC mediates transport of the light chain from the endocytic vesicle
CC to the cytosol. After translocation, the light chain (L)
CC hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC neurotransmitter release. Inhibition of acetylcholine release
CC results in flaccid paralysis, with frequent heart or respiratory
CC failure (By similarity).
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H) (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X73423; CAA51824.1; -.
DR EMBL; X87974; CAA61234.1; -.
DR PIR; I40645; I40645.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0 0 BY SIMILARITY.
FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 223 223 BY SIMILARITY.
FT METAL 226 226 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 429 453 INTERCHAIN (BY SIMILARITY).
FT DISULFID 1234 1279 BY SIMILARITY.
FT TRANSMEM 626 646 POTENTIAL.
FT TRANSMEM 655 675 POTENTIAL.
SQ SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

Query Match 31.1%; Score 790.5; DB 1; Length 1295;
Best Local Similarity 38.8%; Pred. No. 4.5e-37;
Matches 175; Conservative 89; Mismatches 160; Indels 27; Gaps 12;

```
Qy         30 LIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN--DKNQFKLTSSAD 87

Db        861 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 920

Qy         88 SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISI 147

Db        921 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSL 977

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLESN 206

Db        978 NYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQ 1037

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266

Db       1038 KPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1097

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILI---RSKYNQNSNYINYRNLYIG 323

Db       1098 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLN-STLYEG 1156

Qy        324 EKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI-ISDS 382

Db       1157 TKFIIKKYASGN--EDNIVRNNDRVYINVVVKNKEYRL-ATNASQAGVEKILSALEIPDV 1213

Qy        383 NEFYKTIEIKEYDEQP-SYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441

Db       1214 GNLSQVVVMKSKDDQGIRNKCKM--NLQDNNGNDIGFIGFHLYDNIAKL--------VAS 1263

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472

Db       1264 ASNWYNRQIERSSRTLGCSWEFIPVDDGWGE 1294
```

```
Db       1264 NWYNRQVGKA--SRTFGCSWEFIPVDDGWGE 1292


RESULT 5
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             29.6%;  Score 752;  DB 1;  Length 1274;
  Best Local Similarity   37.7%;  Pred. No. 6.3e-35;
  Matches  171;  Conservative   93;  Mismatches  152;  Indels   38;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  ||:   :| :: ||::|| :|  ||:||||: : :   | :   ::||| : :
Db        851 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 910

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCM-KNNSGW 143
              |  |:: : || :||:|| : :||:|||:||||:      |    |||||||  |||||
Db        911 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGW 967

Qy        144 KISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYI 199
              |||:|  |   ||||| | :|  ::: | |     || |||:| ||||||| | |::|||
Db        968 KISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYI 1027

Qy        200 NGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQ 259
              || |     | ::|:: |:  | ||:  |  | | :: ::|| :|||:|::: |: :|  :
Db       1028 NGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNE 1086

Qy        260 SYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRN 319
                ||::||| |:|||:||::||   |: || |  :| :  |  :   : | ::||:
Db       1087 PDPSILKNYWGNYLLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK- 1143

Qy        320 LYIGEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSI 378
              || | : |||:       | |: ||| |  ::::|    |:|:||    | ::||:  :
Db       1144 LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIRT- 1200

Qy        379 ISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYF 438
               |: |:    | :    :    :| : |: :  |  :|||:| |              :
Db       1201 -SNLNDSLGQIIVM---DSIGNNCTMNFQNNNGS--NIGLLGFH-----------SNNL 1242

Qy        439 CISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
               | ||   ::|    |: || |  | |: || |
Db       1243 VASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 1274


RESULT 6
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
```

```
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;
```

```
  Query Match             28.5%;  Score 723;  DB 1;  Length 1250;
  Best Local Similarity   35.3%;  Pred. No. 2.6e-33;
  Matches  159;  Conservative   99;  Mismatches  155;  Indels   38;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:||:::  :| ||| : : :   |     :|||| : :
Db        832 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 891

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
                 |:: ::||  ||::: : :||:|||:||| | |  :   ::||||||||||: |||||
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 949

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |:|:  | ||||  |  |  : : | |     ||:|||:| ||||||: | ::|:||||
Db        950 KVSLNHNEIIWTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::| : |:  | ||:  :   |::| ::||:||: :|::: |: :|  :  :
Db       1010 LIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1068

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
                |||||| |:|:||||: |    |::|   |||::       |  |  |  :    ||
Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTL------SINNIRSTILLANRLYS 1122

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEE-WRVYAYKYFKEQEEKLFLSIISD 381
              | |  |:| :|| | ||::||| | :::: |      : :||      :|: : :|  |
```

```
Db       1123 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKIS--SS 1179

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              | | : : :         :| : || :  : ::|||:|               |    |
Db       1180 GNRFNQVVVMNSVG-----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1220

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||   :  : : :: || | || :: || |
Db       1221 TWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1249


RESULT 7
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             27.9%;  Score 708;  DB 1;  Length 1250;
  Best Local Similarity   35.7%;  Pred. No. 1.8e-32;
  Matches  161;  Conservative   95;  Mismatches  157;  Indels   38;  Gaps   14;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:||:::  :| ||| : : :   |     :|||| : :
Db        832 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 891

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
                 |:: ::||  ||::: : :||:|||:||| | |  :   ::|||||||||: |||||
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 949

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |:|:  | ||||| | :|  : : | |      ||:|||:| ||||||: | ::|:|||
Db        950 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::| : |:  | ||:  :   |::| ::||:||: :|::: |: :|  :  :
Db       1010 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1068

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              |||||||  |:|:||||: |    |::|    ||::       |  |  | :    ||
Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTL------SINNIRSTILLANRLYS 1122

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEE-WRVYAYKYFKEQEEKLFLSIISD 381
              | |  |:| :|| | ||::||| | :::: |        :||      :|: : :|  |
Db       1123 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKIS--SS 1179

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              | | : : :         :| : || :  : ::|||:|               |    |
Db       1180 GNRFNQVVVMNSVG-----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1220

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||   : |    || |  | || :: || |
Db       1221 TWYYTHM-RDNTNSN-GFFWNFISEEHGWQE 1249


RESULT 8
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
```

```
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             24.2%;  Score 614.5;  DB 1;  Length 1276;
  Best Local Similarity   32.9%;  Pred. No. 3.1e-27;
  Matches  157;  Conservative   89;  Mismatches  150;  Indels   81;  Gaps   16;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
              :::| :: |:| : | :: ||:|: : | |:| ||| |:| | | |:||    | |||:|
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWK 144
              | | || |  | ||:::::: : ||||||:| |    |: |  |||||||| :: |||||
Db        910 SGD-KIIVNLNNNILYSAIYENSSVSFWIKISK----DLTN-SHNEYTIINSIEQNSGWK 963

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTL 203
              : ||   | | | |:| | ||: |:|:     : | |:|||||||| :   |:|||| |
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        204 ESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSE 263
              : :  |:|: || ::  | | :| ::|  | :|:: |:||: :|:  :|  :|: |
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIG 323
              :||:||||| :: |||: |    : ||     :    :|:  :|   |       || |
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVQYPDRS------KLYTG 1132

Qy        324 EKFIIRRESN----SQSINDDIVRKEDYIHLDLVLHHEEWRVY-AYKYFKEQEEKLFLSI 378
                 |:  |:     |: :| |           :::||     :| : ||           |
Db       1133 NPITIKSVSDKNPYSRILNGD----------NIILH----MLYNSRKYM----------I 1168

Qy        379 ISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKK----- 433
              | |::  | | :  |  :   |: :|          ::|  ||  |    :: |
Db       1169 IRDTDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQ 1219

Qy        434 --------------YKDYFCISKWYLKEVKRKPYKSNL---GCNWQFIPKDEGWTE 472
                           || :    |     |:: |       |:|| :| || |
Db       1220 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 9
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
```

```
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
```

```
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190

  Query Match           23.9%;  Score 607;  DB 1;  Length 1314;
  Best Local Similarity  32.2%;  Pred. No. 8.5e-27;
  Matches  142;  Conservative   88;  Mismatches  183;  Indels   28;  Gaps   12;

Qy         46 ILNLRYRDNNLIDLSGYGAKVEVYDGVKL----NDKNQFKLTSSADSKIRVTQNQNIIFN 101
              ||||   :: : |:||: : |  |   :|    | |    | ::  |:: | :  :| :|
Db        885 ILNLDINNDIISDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 943

Qy        102 SMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
               || :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :|||
Db        944 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 1003

Qy        157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGE 214
              |  |: : : |  :: :  : |: |:| |:||||: | :| :|||| |  : :|  :|
Db       1004 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 1062

Qy        215 VIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLM 274
              :  :  || |||   :  |:: :  | ||   ||   |:::|       :|:|||||
Db       1063 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 1122

Qy        275 YNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNS 334
              |: |||:    : :  ::|   :    :     |       | || || | ||||:| : :
Db       1123 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 1182

Qy        335 QSINDDIVRKEDYIHLDLVLHHEEWRVYAY----KYFKEQEEKLFLSIISDSNEFYKTIE 390
               | |  |:  |:| | :  :: |  :  |       |   :  | :   :     || :|
Db       1183 NEI-DSFVKSGDFIKLYVSYNNNE-HIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 1240

Qy        391 IKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKR 450
              :  :  :|| ||     |: ::  :||:| |       :     :|    | ||   :|
Db       1241 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 1295

Qy        451 KPYKSNLGCNWQFIPKDEGWT 471
              |     |||:| |:| |||||
Db       1296 K----ILGCDWYFVPTDEGWT 1312


RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
```

```
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             22.2%;  Score 565;  DB 1;  Length 1290;
  Best Local Similarity   32.4%;  Pred. No. 1.9e-24;
  Matches  156;  Conservative   87;  Mismatches  155;  Indels   84;  Gaps   20;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
              :::| :: |:| : | :: ||:|: | | |:| ||| |:|     |:||      ||| |
Db        853 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 912

Qy         85 SAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 142
              | :   |: ||||:||::|||:  ||:|||||| |:    : |     ||||: :|||||
Db        913 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSG 966

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 201
              | | |   | :::||       :|: | |:|  :    | |:|||||:||| : | ||||||
Db        967 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1025

Qy        202 TLESNMDIKDIGEVIVNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQSNI 252
                |   : :|::  :  :  ||||::          | |    :|:: | ||  :|:  :|
Db       1026 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDI 1084

Qy        253 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNS 312
                ::    |:  :||:||| | |||||| |    | |:     ::  :|:  :: | |
Db       1085 NILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN- 1139

Qy        313 NYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEE 372
                  :   | | ||:|     : ||  ||  | :: |: :::      ||  | : |
Db       1140 ------DFNEGYKIIIKRIRG--NTNDTRVRGGDILYFDMTINNK-----AYNLFMKNE- 1185

Qy        373 KLFLSIISD--SNEFYKTIEIKE----------YDEQPS----YSCQLLFKK--DEESTD 414
                :: :|  | |    | ::|          :  ||     |    :||   : |:
Db       1186 ----TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 1241

Qy        415 DIGLIGIHRFYESGVLRKKYKDYFCISKWY-----LKEVKRKPYKSNL---GCNWQFIPK 466
                | || :||   |              ||     :  ||:  | | |     :| |:|
Db       1242 GICSIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPV 1288

Qy        467 DE 468
               |
Db       1289 SE 1290


RESULT 11
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match              9.5%;  Score 241;  DB 1;  Length 1162;
  Best Local Similarity   22.6%;  Pred. No. 2.3e-06;
  Matches  105;  Conservative   76;  Mismatches  152;  Indels  132;  Gaps   22;

Qy         33 MFNKYNSEILNN-------IILNLRYRDNNLI-DLSGYGAKVEVYDGVKL---NDKNQFK 81
              :|| |   ::          :   : :|||:| | ||     ||    : |    : |
Db        784 LFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGKNTLVEYPKDIGLVYGINNNAIH 843

Qy         82 LTSSADSKIRVTQNQNIIF------NSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIIN 135
              || :         |||| |      | :  :||: ||:|      :  || | ::  :|
Db        844 LTGA---------NQNIKFTNDYFENGLTNNFSIYFWLR------NLNQNTIKSK--LIG 886

Qy        136 CMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR-WFFVTITNN--L 192
                ::| ||:|    | ::: :|| ||  |:::        :|   |: | :: |: |
Db        887 SKEDNCGWEIYFENNGLVFNIIDSNGNEKNIY--------LSNISNKSWHYIVISINRLK 938

Qy        193 DNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNI 252
              |   |:|:  | :| |||:|  :  :  |:   | :      ::::  |: |  :| : |
Db        939 DQLLIFIDNILVANEDIKEILNIYSSDIISLLSDNN-----NVYIEGLSVLNKTINSNEI 993
```

```
Qy        253 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFN----------AGNKNSYIKLVKDSSVGE 301
Db        994 LTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEIAINKIEQNNNIYLS--------- 1044

Qy        302 ILIRSKYNQNSNYINYRNLYIGEKF-IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360
Db       1045 -------NNNENSLNFKPL----KFKLLNTNPNKQYV------------------QKWD 1074

Qy        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEEST----DDI 416
Db       1075 EVIFSVLDGTEKYLDISI--DNNRI----------------QLVDNKNNAKTFIINNDI 1115

Qy        417 GLIGIHRFYESGV---LRKKYKDYFCISKWYLKEVKRK-PYKSNL 457
Db       1116 FISNCLTLTYNNVNVYLSIKNQDY----NWVICDLNHDIPKKSYL 1156


RESULT 12
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;

  Query Match             9.5%;  Score 240.5;  DB 1;  Length 1162;
  Best Local Similarity   22.7%;  Pred. No. 2.4e-06;
  Matches   90;  Conservative   77;  Mismatches  128;  Indels  101;  Gaps   17;

Qy         33 MFNKYNSEILNN-------IILNLRYRDNNLI-DLSGYGAKVEVYDGVKL---NDKNQFK 81
Db        784 LFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGKNTLVEYPKDIGLVYGINNNAIH 843

Qy         82 LTSSADSKIRVTQNQNIIF------NSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIIN 135
Db        844 LTGA---------NQNIKFTNDYFENGLTNNFSIYFWLRNLK------QNTIKSK--LIG 886

Qy        136 CMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR-WFFVTITNN--L 192
Db        887 SKEDNCGWEIYFENDGLVFNIIDSNGNEKNIY--------LSNISNKSWHYIVISINRLK 938

Qy        193 DNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNI 252
Db        939 DQLLIFIDNILVANEDIKEILNIYSSDIISLLSDNN-----NVYIEGLSVLNKTINSNEI 993

Qy        253 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFN----------AGNKNSYIKLVKDSSVGE 301
Db        994 LTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEIAINKIEQNNNIYLSINNENNLNF 1053

Qy        302 ILIRSKY---NQNSNYIN------YRNLYIGEKF----------IIRRESNSQ--SIND 339
Db       1054 KPLKFKLLNTNPNKQYVQKWDEVIFSVLDGTEKYLDISTTNNRIQLVDNKNNAQIFIINN 1113

Qy        340 DI-----------------VRKEDYIHLDLVLHHE 357
Db       1114 DIFISNCLTLTYNNVNVYLSIKNQDYNWVICDLNHD 1149


RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             8.5%;  Score 215.5;  DB 1;  Length 1196;
  Best Local Similarity   21.1%;  Pred. No. 6.3e-05;
  Matches   93;  Conservative   86;  Mismatches  146;  Indels  115;  Gaps   22;

Qy         24 SMADTILIEMFNKYNSEILNNIILNLRYRDNNLI-DLSG------YGAKVEVYDGVKLND 76
Db        822 SQVEQVMKEILSPYQL-----LLFASKGPNSNIIEDISGKNTLIQYTESIELVYGV---- 872

Qy         77 KNQFKLTSSADSKIRVTQNQNIIFNSMFL------DFSVSFWIRIPKYRNDDIQNYIHNE 130
Db        873 --------NGESLYLKSPNETIKFSNKFFTNGLTNNFTICFWLRFTG-KNDDKTRLIGNK 923

Qy        131 YTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYIN-RWFFVTIT 189
```

```
                 || ||:|    | ::: :|| ||  :||:        :|  ||  |::::|:
Db        924 V-------NNCGWEIYFEDNGLVFEIIDSNGNQESVY--------LSNIINDNWYYISIS 968

Qy        190 NN--LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQL 247
              :   |   |:||    :|: |  |  :     |:       |::   |:::  |: :  :
Db        969 VDRLKDQLLIFINDKNVANVSIDQILSIYSTNIISL-----VNKNNSIYVEELSVLDNPI 1023

Qy        248 NQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSK 307
              :   |     : |::|   : | ||| | ::|       |:    ::|:
Db       1024 TSEEVIRNYFSYLDNSYIRDSSKSLLEYNKNYQLYN------YV--FPETSL-------- 1067

Qy        308 YNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYF 367
              |  | |  :| :|            |:  ||   |: :       ::: :| :||  |:
Db       1068 YEVNDNNKSYLSL----------KNTDGINISSVKFK-------LINIDESKVYVQKW- 1108

Qy        368 KEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEE-----STDDIGLIGIH 422
                ::  : :: |  | | |        :   ||:  ||       :||
Db       1109 ----DECIICVL-DGTEKYLDI------SPENNRIQLVSSKDNAKKITVNTD-------- 1149

Qy        423 RFYESGVLRKKYKD-YFCIS 441
              :    :   | | || :|
Db       1150 -LFRPDCITFSYNDKYFSLS 1168


RESULT 14
RMAR_CANGA
ID   RMAR_CANGA     STANDARD;      PRT;   339 AA.
AC   P21358;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Mitochondrial ribosomal protein VAR1.
GN   VAR1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OG   Mitochondrion.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86011564; PubMed=3900417;
RA   Ainley W.M., Macreadie I.G., Butow R.A.;
RT   "VAR1 Gene on the mitochondrial genome of Torulopsis glabrata.";
RL   J. Mol. Biol. 184:565-576(1985).
CC   -!- FUNCTION: Essential for mitochondrial protein synthesis and
CC       required for the maturation of small ribosomal subunits (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE VAR1 FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02893; CAA26652.1; -.
DR   PIR; S04682; S04682.
DR   InterPro; IPR007980; Yeast_VAR1.
DR   Pfam; PF05316; Yeast_VAR1; 1.
KW   Ribosomal protein; Mitochondrion.
SQ   SEQUENCE   339 AA;  40828 MW;  354820A9C1FE9E50 CRC64;

  Query Match             6.2%;  Score 158;  DB 1;  Length 339;
  Best Local Similarity   24.0%;  Pred. No. 0.025;
  Matches   82;  Conservative   56;  Mismatches  119;  Indels   84;  Gaps   17;

Qy         30 LIEMFNKYNS-EILNNI----ILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTS 84
              | |: || || : |||:    : |  |  || |:          |   || :|  :|| |
Db         44 LSELNNKGNSLQHLNNMNNWKLQNYNYNKNNTIN---------NYINSKLINKLLYKLMS 94

Qy         85 SADSKIRVTQ-----NQNIIFNSMFLDF--------SVSFWIR-IPKYRN---------D 121
                ::|| :::      | |:| |  |  :         :::::|  | |  |
Db         95 LKNNKIIISKPLYKINMNVI-NIRFYYYNMNNYNYNNINYYINMINKLMNRLNINMNNLS 153

Qy        122 DIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW---TLIDINGKTKSVFFEY-----NIR 173
              :| :| :|:  || :|      |      | |:    :|:| |    :  ||     ||
Db        154 NILSYYYNKKVIIEPIK----LKYLYNNNEIMTKYISLLDNNKYNNGLLMEYQRTLNNIM 209

Qy        174 EDISEYINRWFFVTITNNLDNAKIYINGTLESNMDIKDI-GEVIVNGEITFKLDGDVDRT 232
                ::::     ::   ||::  | | |  | :| :| :|   : :|  :
Db        210 PKLNDHNISMNYINNINNINKLK-YNNILLNNNNNINNIYNNININNNM---------- 257

Qy        233 QFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIK 292
              :  ||   ::  |     | | :| :       :  |    ||:|     || |:  |
Db        258 NLLMFKYLIGWSIMLKGRLNKNISRIST------TYLNNGTFNNKKYLW---GNLNNNFK 308

Qy        293 LVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN 333
              |          :  | |:|  || |:     |: |: : |
Db        309 L-----------NYINSNNNIYNYNNINKNGKYNIKVKLN 337


RESULT 15
MTLD_MYCPU
ID   MTLD_MYCPU     STANDARD;      PRT;   360 AA.
AC   Q98PH2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17).
GN   MTLD OR MYPU_7500.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
CC   -!- CATALYTIC ACTIVITY: D-mannitol 1-phosphate + NAD(+) = D-fructose
CC       6-phosphate + NADH.
CC   -!- SIMILARITY: Belongs to the mannitol dehydrogenase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL445565; CAC13923.1; -.
DR   PIR; F90605; F90605.
DR   MypuList; MYPU_7500; -.
DR   HAMAP; MF_00196; -; 1.
DR   InterPro; IPR000669; Mannitol_dh.
DR   Pfam; PF01232; Mannitol_dh; 1.
DR   PRINTS; PR00084; MTLDHDRGNASE.
DR   PROSITE; PS00974; MANNITOL_DHGENASE; FALSE_NEG.
KW   Oxidoreductase; NAD; Complete proteome.
FT   NP_BIND       6     17       NAD (BY SIMILARITY).
SQ   SEQUENCE   360 AA;  42409 MW;  E60949DC40914B1D CRC64;

  Query Match             5.9%;  Score 150;  DB 1;  Length 360;
  Best Local Similarity   18.8%;  Pred. No. 0.076;
  Matches   87;  Conservative   71;  Mismatches  131;  Indels  174;  Gaps   19;

Qy         11 HHSSGHIEGRHMAS------------MADTILIEMFNKYNSEILNNIILNLRYRDNNL- 56
                |  :|:| || : |              |  |||  ||  |  : :   |  :: :|
```

```
Db          8 HFGAGNI-GRGLISDIYMKNNMDFALVDIDKDLIEKLNKQKSYSIIDFQTNKVFQISNFK 66

Qy         57 ----------------------IDLSGYGAKVEVYDGVKLNDK-------NQFKLTSSA 86
                                    |  |    :  : :: |||  :|        | :|::|
Db         67 AFSIDQEDEIKKWIEQADFISTSIGWSNLASLKKFFENVKLKEKAQIICFENGYKISSFF 126

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
               | : :  |        |:: ||     || :::| :  |: : | ||   || |  |::
Db        127 QSILNIDSNH-------FVNASVD--KIIPNFKSDSLDVYVESYYEIILEQKNESQKKLN 177

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
                                    |    |:  |||:  |  : | : :   |:
Db        178 F----------------------VNYSTDLEAYINKKLF--LVNAIHSTIGYLG------ 207

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
                                          ::| ::  |  ||   |  ::||:  :: :
Db        208 ----------------------------YLKKYTYINEALNDQQI--LFKIKRLAKIIN 236

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYI-KLVKDSSVGE---ILIR------SKYNQNSNYIN 316
               :      : :||| :|     | |: | :|  |: |   :: |       | ::|  |
Db        237 E------ILSKEYLLFKVDYLNDYLEKNLKRFSIKENQDLISRVARNPIQKLSKNERYFL 290

Qy        317 YRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYK---YFKEQEEK 373
              ||                         |:  : :|::|          ||   |:  : :|
Db        291 IYNLV--------------------KKHNLEIDILLE-------IYKSIFYYDNKMDK 321

Qy        374 LFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDI 416
                :|::   ||| |     :|: :     |:|  : |
Db        322 -------ESSKIQSTIENKSL----AYALKKFSNLDQEDQEKI 353
```

Search completed: September 10, 2004, 03:35:15
Job time : 9.75902 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 21.5963 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-44
Perfect score: 2540
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YKSNLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2358	92.8	1291	2	Q08077	Q08077 clostridium
2	2145	84.4	1291	2	Q8GR96	Q8gr96 clostridium
3	2119	83.4	1291	2	Q933K0	Q933k0 clostridium
4	2113	83.2	1291	2	Q9ZAJ8	Q9zaj8 clostridium
5	2108	83.0	1291	2	Q93G71	Q93g71 clostridium
6	2090	82.3	441	2	Q9X708	Q9x708 clostridium
7	798	31.4	1268	2	Q45851	Q45851 clostridium
8	772	30.4	1278	2	Q57236	Q57236 clostridium
9	766.5	30.2	1280	2	Q9ZAJ5	Q9zaj5 clostridium
10	764	30.1	361	2	Q45848	Q45848 clostridium
11	764	30.1	361	2	Q45846	Q45846 clostridium
12	721	28.4	1251	2	Q9K395	Q9k395 clostridium
13	708.5	27.9	1252	2	Q8KZM3	Q8kzm3 clostridium
14	708.5	27.9	1255	2	Q9FAR6	Q9far6 clostridium
15	609.5	24.0	1275	12	Q9QTG7	Q9qtg7 clostridium
16	606	23.9	451	2	Q9LA13	Q91a13 clostridium
17	604.5	23.8	1280	2	Q45849	Q45849 clostridium
18	604.5	23.8	1280	2	Q9LBS7	Q9lbs7 clostridium
19	604.5	23.8	1280	2	Q841S3	Q841s3 clostridium
20	579	22.8	1310	2	Q93N27	Q93n27 clostridium
21	565	22.2	1291	2	Q93HT3	Q93ht3 clostridium
22	563	22.2	1285	2	Q45967	Q45967 clostridium
23	562	22.1	1285	2	Q9LBR1	Q9lbr1 clostridium
24	368	14.5	78	2	Q7WUH9	Q7wuh9 clostridium
25	366	14.4	78	2	Q7WRW0	Q7wrw0 clostridium
26	324.5	12.8	367	2	Q45861	Q45861 clostridium
27	324.5	12.8	367	2	Q45862	Q45862 clostridium
28	235.5	9.3	1198	2	O06018	O06018 clostridium
29	227.5	9.0	1198	2	O69277	O69277 clostridium
30	226.5	8.9	1198	2	Q45893	Q45893 clostridium
31	222	8.7	1160	2	O87710	O87710 clostridium
32	219	8.6	1161	2	Q45891	Q45891 clostridium
33	218.5	8.6	1196	2	Q9LBS8	Q9lbs8 clostridium
34	215.5	8.5	1196	2	Q45916	Q45916 clostridium
35	215.5	8.5	1196	2	Q93HT4	Q93ht4 clostridium
36	215.5	8.5	1196	9	Q38197	Q38197 clostridium
37	214	8.4	1161	2	O69276	O69276 clostridium
38	214	8.4	1193	2	Q45880	Q45880 clostridium
39	214	8.4	1197	2	Q45888	Q45888 clostridium
40	213.5	8.4	1196	2	Q9LBR2	Q9lbr2 clostridium
41	212	8.3	1197	2	Q9ZAJ9	Q9zaj9 clostridium
42	212	8.3	1197	2	P71117	P71117 clostridium
43	212	8.3	1197	2	O33871	O33871 clostridium
44	209.5	8.2	1162	2	Q9ZAJ6	Q9zaj6 clostridium
45	208.5	8.2	1196	2	Q53550	Q53550 clostridium

ALIGNMENTS

RESULT 1
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR PIR; I40631; I40631.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match           92.8%;  Score 2358;  DB 2;  Length 1291;
  Best Local Similarity   99.5%;  Pred. No. 4.9e-116;
  Matches  442;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 2
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match           84.4%;  Score 2145;  DB 2;  Length 1291;
  Best Local Similarity   89.2%;  Pred. No. 7.7e-105;
  Matches  398;  Conservative   25;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              :|||||:|||||||||||||||||||||||||||||| |||||||:||||||||||||
Db        846 NTILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELNDKNQFKLTSST 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|:|||||||||||||||||||||||||||||:|| ||||||||||||||:|||||||||
Db        906 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||||||||| |||||||||||:||||||:|||||||||||||| |||||||| ||||
Db        966 IRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNSDNAKIYINGKLESN 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              :||||||||| |||| |||||||:|||||||||||||||:|:|||||||||||||||||
Db       1026 IDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||||||| |||||||| |||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    : |||||||| |||:|:||||: | ||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLFLANIYDSNEFY 1205

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              |||:|||||||:|||||||||||||||:|||||||||||||:: | ||:|||||||||||
Db       1206 KTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFCISKWYLK 1265

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              |||||||  |||||||||||||||| |
Db       1266 EVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 3
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
```

```
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             83.4%;  Score 2119;  DB 2;  Length 1291;
  Best Local Similarity   88.1%;  Pred. No. 1.8e-103;
  Matches  393;  Conservative   30;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              :|||||:|||||||:|||||||||||||| |||||||||||||||||||||||||||||||
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|||||||||||||||||||||||||||||:|| |||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||| |||||||||||| |||||||:|:|||||||||||||||| ||||||||| |||:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNSDNAKIYINGKLESH 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              :||:|| ||| | || |||||::||||||||||||||:|:||||:|||||||||||||||
Db       1026 IDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||||| |||||||| |||||||| |||||:|||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:||||||||||||||:||    ::||||| |||||::|||||: ||||:|||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              ||:||||||||:|||||||||||||||:||||||||||:: |:||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||| | ||||||||||||||||
Db       1266 EVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 4
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match             83.2%;  Score 2113;  DB 2;  Length 1291;
  Best Local Similarity   87.9%;  Pred. No. 3.7e-103;
  Matches  392;  Conservative   30;  Mismatches   24;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              :|||||:|||||||:|||||||||||||| |||||||||||||||||||||||||||||||
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :||||| |||||||||||||||||||||:|| |||||||||||||||||||||||||||||
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||| |||||||||||| |||||||:|:|||||||||||||||| ||||||||| |||:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNSDNAKIYINGKLESH 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              :||:|| ||| | || |||||::||||||||||||||:|:||||:|||||||||||||||
Db       1026 IDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||||| |||||||| |||||||| |||||:|||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:||||||||||||||:||    ::||||| |||||::|||||: ||||:|||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              ||:||||||||:|||||||||||||||:||||||||||:: |:||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||| | ||||||||||||||||
Db       1266 EVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 5
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             83.0%;  Score 2108;  DB 2;  Length 1291;
  Best Local Similarity   87.7%;  Pred. No. 6.8e-103;
  Matches  391;  Conservative   30;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              :|||||:||||||:|||||||||||||| ||||||||||||||||||||||||||||||
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 905

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :||||| ||||||||||||||||||||||:|| ||||||||||||||||||||||||||
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||| |||||||||| |||||||:|:|||||||||||||||| ||||||||  |||:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNSDNAKIYINGKLESH 1025

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              :||:|| ||| | || |||||::|||||||||||||:|:||||:|||||||||||||||
Db       1026 IDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    ::||||| |||||::||||||: ||||:|||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYIYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              ||:|||||||:|||||||||||||||:|||||||||||:: |:||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||||||| | ||||||||||||||||
Db       1266 EVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 6
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;

  Query Match             82.3%;  Score 2090;  DB 2;  Length 441;
  Best Local Similarity   88.8%;  Pred. No. 1.9e-102;
  Matches  389;  Conservative   26;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 86
              :|||||:||||||:|||||||||||||| ||||||||||||||||||||||||||||||
Db          4 NTILIEIFNKYNSKILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELNDKNQFKLTSST 63

Qy         87 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 146
              :|:|||||||||||||||||||||||||||:|| ||||||||||||:|||||||||
Db         64 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKIS 123

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESN 206
              |||||||||| |||||||||||:||||||:|||||||||||| |||||||| ||||
Db        124 IRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNSDNAKIYINGKLESN 183

Qy        207 MDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLK 266
              :||||||||| |||| ||||||:||||||||||||:|:||||||||||||||||||||
Db        184 IDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYLK 243

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKF 326
              |||||||||||||||||||||||||||||| |||||||| |||||||||||||||||||||
Db        244 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEKF 303

Qy        327 IIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFY 386
              ||||:|||||||||||||:||    : |||||||| |||:|:|| |: | ||||||
Db        304 IIRRKSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLVLANIYDSNEFY 363

Qy        387 KTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLK 446
              |||:||||||:|||||||||||||:|||||||||||:: | ||:||||||||||
Db        364 KTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFCISKWYLK 423

Qy        447 EVKRKPYKSNLGCNWQFI 464
              ||||||| |||||||||
Db        424 EVKRKPYNPNLGCNWQFI 441


RESULT 7
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
```

```
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match           31.4%;  Score 798;  DB 2;  Length 1268;
  Best Local Similarity  38.6%;  Pred. No. 4.9e-34;
  Matches  174;  Conservative   95;  Mismatches  152;  Indels   30;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | ||| :  ::   |::: |||::| :|  || ||||: : :   : :   ::||| : |
Db        843 DKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYS 902

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: :|||  ||:|| : :|||||:|||| |  ::| ::|||||||||: |||||
Db        903 SRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN--LKN-LNNEYTIINCMRNNNSGW 959

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |||:  | ||||| |  |  : : | |    |||:|||:| ||||||| | ::|:||||
Db        960 KISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGN 1019

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::| : |:  | ||: |  | |::: ::|| ||| :|::: |: :|  :  |
Db       1020 LTDQKSILNLGNIHVDDNILFKIVGCND-TRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              |||||||  |:|||:||: |    |  : : |:|  :  |  :      :|  :   ||
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPN--MSVTKNSDILNINRQRGIYSKTNIFSNARLYT 1136

Qy        323 GEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISD 381
              | : |||:  :: : | |: ||| | :::::|  : |:::||       |: : | ||:
Db       1137 GVEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRISN 1196

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              ||  | : ::   |    :| : || :  : :||||:| |            :    |
Db       1197 SN--YNSNQMIIMDSIGD-NCTMNFKTN--NGNDIGLLGFH-----------LNNLVAS 1239

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
               || | ::      | || | || |: || |
Db       1240 SWYYKNIRNN--TRNNGCFWSFISKEHGWQE 1268


RESULT 8
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           30.4%;  Score 772;  DB 2;  Length 1278;
  Best Local Similarity  37.1%;  Pred. No. 1.1e-32;
  Matches  173;  Conservative   87;  Mismatches  146;  Indels   60;  Gaps   14;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||   :| :| ||::|| :|  ||:|||: : :   | :   ::||| : |
Db        852 DKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYS 911

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: : || :||:|  : :||:|||:||||| |   :  ::||||||:|:: |||||
Db        912 SKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFN---KVNLNNEYTIIDCIRNNNSGW 968

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |||:  |:|||||  |  |  : : | |     ||:|||:| |||||||  | |::|||||
Db        969 KISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGN 1028

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::|:: |:  | ||: |  | |::: ::|| :|:|:| :: |: :|  :
Db       1029 LIDEKSISNLGDIHVSDNILFKIVGCND-TRYVGIRYFKVFDTELGKTEIETLYSDEPDP 1087

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYR---- 318
              |||||||  |:||| ||: |          |  | |:           ||||::|
Db       1088 SILKDFWGNYLLYNKRYYLLNL--------LRTDKSI---------TQNSNFLNINQQRG 1130

Qy        319 -----------NLYIGEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKY 366
                        || | : |||:  ::   | |: ||| |  ::::|    |:|:||
Db       1131 VYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADIS 1190

Qy        367 FKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYE 426
              : |: : |   |:||     | :    :    :| : |: :  :  :|||:| |
Db       1191 IAKPEKIIKLIRTSNSNNSLGQIIVM---DSIGNNCTMNFQNN--NGGNIGLLGFH---- 1241

Qy        427 SGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
                    :    | ||   : ||   || || | || |: || |
Db       1242 --------SNNLVASSWYYNNI-RKNTSSN-GCFWSFISKEHGWQE 1277
```

```
RESULT 9
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             30.2%;  Score 766.5;  DB 2;  Length 1280;
  Best Local Similarity   37.7%;  Pred. No. 2.2e-32;
  Matches  174;  Conservative   91;  Mismatches  149;  Indels   47;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  ||:   :| :: ||::|| :|  ||:||||: : :   | :   ::||| : |
Db        852 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYS 911

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCM-KNNSGW 143
                |:: : || :||:|| : :||:|||: |||:     :|    ||||||||  |||||
Db        912 GRLSEVNIAQNNDIIYNSRYQNFSISFWVTIPKHYRPMNRN---REYTIINCMGNNNSGW 968

Qy        144 KISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYI 199
              |||:|  |   ||||| | :|  : : | |      ||:|||:| |||||| | |::|||
Db        969 KISLRTIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGNSRIYI 1028

Qy        200 NGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQ 259
              || |     | ::|:: |:  | ||: |  | | :: ::|| :|||:|::: |: :|  :
Db       1029 NGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNE 1087

Qy        260 SYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQN-------S 312
                 |||:||| |:|||:||:||   |: ||   ::| :  |      ||          |
Db       1088 PDPSILKDYWGNYLLYNKKYYLFNLLRKDKYI--TRNSGILNI------NQQRGVTGGIS 1139

Qy        313 NYINYRNLYIGEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQE 371
              ::||: || | : |||: :     | |: ||| |  ::::| |  |:|:||     : |
Db       1140 VFLNYK-LYEGVEVIIRKNAPIDISNTDNFVRKNDLAYINVVDHGVEYRLYADISITKSE 1198

Qy        372 EKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLR 431
              : : |   |: |:    | :    :    :| : |: :: |  :|||:| |
Db       1199 KIIKLIRTSNPNDSLGQIIVM---DSIGNNCTMNFQNNDGS--NIGLLGFH--------- 1244

Qy        432 KKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
                  |    | ||   ::|    |: || | || |: || |
Db       1245 ---SDDLVASSWYYNHIRRN--TSSNGCFWSFISKEHGWKE 1280


RESULT 10
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;

  Query Match             30.1%;  Score 764;  DB 2;  Length 361;
  Best Local Similarity   98.6%;  Pred. No. 7.8e-33;
  Matches  145;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 274

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 334

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||||||||
Db        335 GNRIIWTLIDINGKTKSVFFEYNIRED 361


RESULT 11
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
```

```
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;

  Query Match             30.1%;  Score 764;  DB 2;  Length 361;
  Best Local Similarity   98.6%;  Pred. No. 7.8e-33;
  Matches  145;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 88
              |||::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 274

Qy         89 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 334

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||||||
Db        335 GNRIIWTLIDINGKTKSVFFEYNIRED 361


RESULT 12
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;
```

```
  Query Match           28.4%;  Score 721;  DB 2;  Length 1251;
  Best Local Similarity   34.9%;  Pred. No. 5.4e-30;
  Matches  157;  Conservative  100;  Mismatches  157;  Indels   36;  Gaps   12;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:||:::  :| ||| : : :   : :   :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFN 892

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: ::||  ||::: : :||:|||:||| | |  :   |:|||||||||: |||||
Db        893 SKPSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--INNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |:|:  | ||||| |     : : |:|     ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGH 1010

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::| : |:  | ||:  :   |::| ::||:|: :|::: |: :|  :  :
Db       1011 LIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              ||||||| |:|:| ||: |    |::|   |||::      |  |  |  :    ||
Db       1070 NILKDFWGNYLLYDKGYYLLNVLKPNNFIDRRKDSTL------SINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDS 382
              | |  |:|  |  | ||  ||| | :::: : :   : :||     ::|: :  |  |
Db       1124 GIKVKIQR-VNDSSTNDRFVRKNDQVYINYISNSSSYSLYADTNTTDKEKTIKSS--SSG 1180

Qy        383 NEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISK 442
              | | : : :          :| : || :  : ::|||:|          |    |
Db       1181 NRFNQVVVMNSVGN----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAST 1222

Qy        443 WYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||   :  : : :: || | || :: || |
Db       1223 WYYTHM--RDHTNSNGCFWNFISEEHGWQE 1250
```

```
RESULT 13
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
```

```
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match           27.9%;  Score 708.5;  DB 2;  Length 1252;
  Best Local Similarity   35.7%;  Pred. No. 2.4e-29;
  Matches  161;  Conservative   95;  Mismatches  158;  Indels   37;  Gaps   14;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:||:::  :| ||| : : :   |     :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
              |:: ::||  ||::: : :||:|||:||| | |  :   ::||||||||: |||||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
              |:|:  | ||||| | :|  : : | |     ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
              |      | ::| : |:  | ||:  :   |::| ::||:|: :|::: |: :|  :  :
Db       1011 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
              ||||||| |:|:|||| : |    |::|    ||::      |  |  |  :    ||
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTL------SINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEE-WRVYAYKYFKEQEEKLFLSIISD 381
              | |  |:| :|| | ||::||| | :::: |         :||      :|: : :|  |
Db       1124 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKIS--SS 1180

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
              | | : : :          :| : || :  : ::|||:|                |    |
Db       1181 GNRFNQVVVMNSVGN----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1222

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
              ||   : |    || | | || :: || |
Db       1223 TWYYTHM-RDNTNSN-GFFWNFISEEHGWQE 1251
```

```
RESULT 14
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
```

```
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match           27.9%;  Score 708.5;  DB 2;  Length 1255;
  Best Local Similarity  35.7%;  Pred. No. 2.5e-29;
  Matches  161;  Conservative   95;  Mismatches  158;  Indels   37;  Gaps   14;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKLTS 84
Db        836 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 895

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMK-NNSGW 143
Db        896 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 953

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGT 202
Db        954 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1013

Qy        203 LESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYS 262
Db       1014 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1072

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYI 322
Db       1073 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTL------SINNIRSTILLANRLYS 1126

Qy        323 GEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEE-WRVYAYKYFKEQEEKLFLSIISD 381
Db       1127 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKIS--SS 1183

Qy        382 SNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCIS 441
Db       1184 GNRFNQVVVMNSVGN----NCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1225

Qy        442 KWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
Db       1226 TWYYTHM-RDNTNSN-GFFWNFISEEHGWQE 1254


RESULT 15
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match           24.0%;  Score 609.5;  DB 12;  Length 1275;
  Best Local Similarity  32.8%;  Pred. No. 4e-24;
  Matches  156;  Conservative   89;  Mismatches  149;  Indels   81;  Gaps   16;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 84
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         85 SADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWK 144
Db        910 SGD-KIIVNLNNNILYSAIYENSSVSFWIKISK----DLTN-SHNEYTIINSIEQNSGWK 963

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTL 203
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        204 ESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSE 263
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIG 323
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVQYPDRS------KLYTG 1132

Qy        324 EKFIIRRESN----SQSINDDIVRKEDYIHLDLVLHHEEWRVY-AYKYFKEQEEKLFLSI 378
Db       1133 NPITIKSVSDKNPYSRILNGD----------NIILH----MLYNSRKYM----------I 1168

Qy        379 ISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKK----- 433
Db       1169 IRDTDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQ 1219

Qy        434 --------------YKDYFCISKWYLKEVKRKPYKSNL---GCNWQFIPKDEGW 470
Db       1220 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 1274


Search completed: September 10, 2004, 03:41:27
Job time : 26.5963 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 27.6553 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	472	100.0	472	2	AAW68394	Aaw68394 Clostridi
2	446	94.5	1070	3	AAY93308	Aay93308 A mangane
3	446	94.5	1095	3	AAY93311	Aay93311 A mangane
4	368	78.0	1291	2	AAW68392	Aaw68392 Clostridi
5	350	74.2	440	3	AAY77135	Aay77135 Synthetic
6	350	74.2	440	4	AAB04091	Aab04091 Botulism
7	324	68.6	1169	2	AAW56017	Aaw56017 Recombina
8	224	47.5	439	4	AAB04085	Aab04085 Botulism
9	140	29.7	848	4	AAB04082	Aab04082 Botulism
10	70	14.8	472	2	AAW68393	Aaw68393 Clostridi
11	31	6.6	473	2	AAW68400	Aaw68400 Clostridi
12	26	5.5	448	2	AAW68399	Aaw68399 Clostridi
13	26	5.5	451	2	AAW68395	Aaw68395 Clostridi
14	26	5.5	451	2	AAW68398	Aaw68398 Clostridi
15	26	5.5	452	2	AAW68396	Aaw68396 Clostridi
16	26	5.5	462	2	AAR95009	Aar95009 Type A ne
17	26	5.5	462	2	AAW68390	Aaw68390 Clostridi
18	26	5.5	462	2	AAW68397	Aaw68397 Clostridi
19	23	4.9	393	4	AAG68149	Aag68149 Streptomy
20	23	4.9	393	4	AAB74844	Aab74844 S. griseu
21	23	4.9	423	2	AAW68098	Aaw68098 Chlamydom
22	22	4.7	22	4	AAB67674	Aab67674 Amino aci
23	22	4.7	22	5	AAM49518	Aam49518 His-tag p
24	22	4.7	26	7	ADD31745	Add31745 Purificat
25	22	4.7	53	2	AAW36986	Aaw36986 HFKT pept
26	22	4.7	54	2	AAW36990	Aaw36990 HFKT pept
27	22	4.7	91	2	AAY09127	Aay09127 Recombina
28	22	4.7	93	2	AAY09128	Aay09128 Recombina
29	22	4.7	195	3	AAB11377	Aab11377 B. stearo
30	22	4.7	269	5	AAM49520	Aam49520 B. mori 1
31	22	4.7	269	5	AAM49541	Aam49541 B. mori T
32	22	4.7	323	2	AAW71185	Aaw71185 FP505 pro
33	22	4.7	332	4	AAE02191	Aae02191 Staphyloc
34	22	4.7	364	3	AAB13011	Aab13011 Q-tagged
35	22	4.7	567	4	AAU00220	Aau00220 Bad-DTTR
36	21	4.4	21	2	AAW23653	Aaw23653 Recombina
37	21	4.4	21	2	AAW23650	Aaw23650 Recombina
38	21	4.4	21	2	AAW23647	Aaw23647 Recombina
39	21	4.4	194	5	AAM51139	Aam51139 HIV TAT t
40	20	4.2	411	4	AAU00219	Aau00219 Bcl-Xl-DT
41	17	3.6	24	2	AAW19779	Aaw19779 Polyhisti
42	17	3.6	24	2	AAY06647	Aay06647 Polyhisti
43	17	3.6	24	2	AAY27015	Aay27015 Met Gly H
44	17	3.6	24	2	AAY09370	Aay09370 12 amino
45	17	3.6	24	3	AAY81959	Aay81959 Enterokin

ALIGNMENTS

RESULT 1
AAW68394
ID AAW68394 standard; protein; 472 AA.
XX
AC AAW68394;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30581.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 303-305; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant

```
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             100.0%;  Score 472;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 2
AAY93308
ID   AAY93308 standard; protein; 1070 AA.
XX
AC   AAY93308;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype B.
XX
OS   Synthetic.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 45-47; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC   be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype B. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a domain
CC   that translocates the SOD of the composition into the neuronal cell.
CC   After translocation, the linker is cleaved to release the SOD. The
CC   composition is useful for treating neuronal diseases caused or augmented
CC   by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC   disease, Huntington's disease and motor neurone diseases
XX
SQ   Sequence 1070 AA;

  Query Match             94.5%;  Score 446;  DB 3;  Length 1070;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 684

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 744

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 804

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 864

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 924

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 984

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1044

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1045 EVKRKPYNLKLGCNWQFIPKDEGWTE 1070


RESULT 3
AAY93311
ID   AAY93311 standard; protein; 1095 AA.
XX
AC   AAY93311;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
```

```
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype B.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 54-57; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide dismutase
CC   (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC   cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC   serotype B. The specification describes a composition for delivery of SOD
CC   to neuronal cells. The composition comprises SOD linked, by a cleavable
CC   linker, to a neuronal cell targeting component (NCTC). This component has
CC   a domain that binds to a neuronal cell and a domain that translocates the
CC   SOD of the composition into the neuronal cell. After translocation, the
CC   linker is cleaved to release the SOD. The composition is useful for
CC   treating neuronal diseases caused or augmented by oxidative stress, such
CC   as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases
XX
SQ   Sequence 1095 AA;

  Query Match             94.5%;  Score 446;  DB 3;  Length 1095;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        650 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 709

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        710 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 769

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        770 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 829

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        830 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 889

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 949

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        950 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1009

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1010 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1069

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1070 EVKRKPYNLKLGCNWQFIPKDEGWTE 1095


RESULT 4
AAW68392
ID   AAW68392 standard; protein; 1291 AA.
XX
AC   AAW68392;
XX
DT   17-OCT-2003  (revised)
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type B toxin.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30579.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 291-296; 428pp; English.
XX
CC   This is the amino acid sequence of the type B toxin of Clostridium
CC   botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC   B toxin has been expressed as a histidine-tagged protein in Escherichia
CC   coli. The invention relates to C. botulinum recombinant toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of the recombinant proteins are E. coli, insect cells and
CC   yeast cells. The recombinant proteins are used as immunogens for the
CC   production of vaccines and antitoxins that are useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin.
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1291 AA;

  Query Match             78.0%;  Score 368;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  368;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEY 394
              ||||||||
Db       1206 NTIQIKEY 1213
```

RESULT 5
AAY77135
ID AAY77135 standard; protein; 440 AA.
XX
AC AAY77135;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype B (BoNTB) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype B; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87213.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 23; Page; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast. Note: This
CC sequence is not given in the specification, but is decoded from the BoNTB
CC Hc DNA sequence given on pages 39-40
XX
SQ Sequence 440 AA;

```
  Query Match             74.2%;  Score 350;  DB 3;  Length 440;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  350;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          35 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQ 94
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           3 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQ 62

Qy          95 NQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW 154
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          63 NQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW 122

Qy         155 TLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIRE 214
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         123 TLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIRE 182

Qy         215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         183 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 242

Qy         275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         243 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 302

Qy         335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDE 384
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         303 QSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDE 352
```

RESULT 6
AAB04091
ID AAB04091 standard; protein; 440 AA.
XX
AC AAB04091;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.

PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54485.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 4b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 440 AA;

Query Match 74.2%; Score 350; DB 4; Length 440;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         35 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQ 94
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQ 62

Qy         95 NQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW 154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 NQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW 122

Qy        155 TLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIRE 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIRE 182

Qy        215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 242

Qy        275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 302

Qy        335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDE 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 QSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDE 352
```

RESULT 7
AAW56017
ID AAW56017 standard; protein; 1169 AA.
XX
AC AAW56017;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH728/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26289.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS Example 2; Page 91-94; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 1169 AA;

Query Match 68.6%; Score 324; DB 2; Length 1169;
Best Local Similarity 100.0%; Pred. No. 4.7e-298;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYL 350
              ||||||||||||||||||||||||
```

```
Db       1146 IIRRKSNSQSINDDIVRKEDYIYL 1169


RESULT 8
AAB04085
ID   AAB04085 standard; protein; 439 AA.
XX
AC   AAB04085;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin C fragment sequence (serotype B).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Example 8; Page 38; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 439 AA;

  Query Match             47.5%;  Score 224;  DB 4;  Length 439;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-203;
  Matches  324;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         34 FNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVT 60

Qy         94 QNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 QNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIT 120

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIR 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIR 180

Qy        214 EVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPL 240

Qy        274 MYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 MYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSN 300

Qy        334 SQSINDDIVRKEDYIYLDFFNLNQE 358
              |||||||||||||||||||||||||
Db        301 SQSINDDIVRKEDYIYLDFFNLNQE 325


RESULT 9
AAB04082
ID   AAB04082 standard; protein; 848 AA.
XX
AC   AAB04082;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain (serotype B).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Disclosure; Page 13; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
```

```
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 848 AA;

  Query Match             29.7%;  Score 140;  DB 4;  Length 848;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-123;
  Matches  140;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        151 RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIK 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIK 587

Qy        211 DIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWG 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 DIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWG 647

Qy        271 NPLMYNKEYYMFNAGNKNSY 290
              ||||||||||||||||||||
Db        648 NPLMYNKEYYMFNAGNKNSY 667


RESULT 10
AAW68393
ID   AAW68393 standard; protein; 472 AA.
XX
AC   AAW68393;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B strain Eklund 17B.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30580.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 300-302; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             14.8%;  Score 70;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-57;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182

Qy        183 WFFVTITNNL 192
              ||||||||||
Db        183 WFFVTITNNL 192


RESULT 11
AAW68400
ID   AAW68400 standard; protein; 473 AA.
XX
AC   AAW68400;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW   botulism; BotG.
XX
OS   Clostridium botulinum; serotype G strain 113/30.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30596.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC   sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC   express BotG soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
```

```
XX
SQ   Sequence 473 AA;

  Query Match             6.6%;  Score 31;  DB 2;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-20;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31
              |||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31


RESULT 12
AAW68399
ID   AAW68399 standard; protein; 448 AA.
XX
AC   AAW68399;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type F toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW   botulism; BotF.
XX
OS   Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 48; Page 364-365; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC   sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC   express BotC soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 448 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 13
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 14
AAW68398
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 23.876 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	472	100.0	472	12	US-10-272-898-46	Sequence 46, Appl
2	472	100.0	472	15	US-10-354-774-46	Sequence 46, Appl
3	472	100.0	472	15	US-10-271-012-46	Sequence 46, Appl
4	472	100.0	472	16	US-10-729-122-46	Sequence 46, Appl
5	472	100.0	472	16	US-10-729-039-46	Sequence 46, Appl
6	446	94.5	1290	15	US-10-360-101-220	Sequence 220, App
7	446	94.5	1290	15	US-10-452-024-119	Sequence 119, App
8	446	94.5	1291	12	US-10-272-898-42	Sequence 42, Appl
9	446	94.5	1291	15	US-10-354-774-42	Sequence 42, Appl
10	446	94.5	1291	15	US-10-271-012-42	Sequence 42, Appl
11	446	94.5	1291	15	US-10-452-024-2	Sequence 2, Appli
12	446	94.5	1291	15	US-10-452-024-118	Sequence 118, App
13	446	94.5	1291	16	US-10-729-122-42	Sequence 42, Appl
14	446	94.5	1291	16	US-10-729-039-42	Sequence 42, Appl
15	350	74.2	440	10	US-09-910-186A-8	Sequence 8, Appli
16	324	68.6	1169	14	US-10-241-596-20	Sequence 20, Appl
17	107	22.7	1291	15	US-10-452-024-122	Sequence 122, App
18	70	14.8	472	12	US-10-272-898-44	Sequence 44, Appl
19	70	14.8	472	15	US-10-354-774-44	Sequence 44, Appl
20	70	14.8	472	15	US-10-271-012-44	Sequence 44, Appl
21	70	14.8	472	16	US-10-729-122-44	Sequence 44, Appl
22	70	14.8	472	16	US-10-729-039-44	Sequence 44, Appl
23	70	14.8	1291	12	US-10-272-898-40	Sequence 40, Appl
24	70	14.8	1291	15	US-10-354-774-40	Sequence 40, Appl
25	70	14.8	1291	15	US-10-271-012-40	Sequence 40, Appl
26	70	14.8	1291	15	US-10-452-024-117	Sequence 117, App
27	70	14.8	1291	15	US-10-205-516-4	Sequence 4, Appli
28	70	14.8	1291	16	US-10-729-122-40	Sequence 40, Appl
29	70	14.8	1291	16	US-10-729-039-40	Sequence 40, Appl
30	70	14.8	1301	15	US-10-205-516-18	Sequence 18, Appl
31	68	14.4	1291	15	US-10-452-024-121	Sequence 121, App
32	68	14.4	1291	15	US-10-452-024-123	Sequence 123, App
33	53	11.2	361	15	US-10-452-024-175	Sequence 175, App
34	53	11.2	361	15	US-10-452-024-176	Sequence 176, App
35	43	9.1	441	15	US-10-452-024-188	Sequence 188, App
36	43	9.1	1291	15	US-10-452-024-142	Sequence 142, App
37	31	6.6	473	12	US-10-272-898-79	Sequence 79, Appl
38	31	6.6	473	15	US-10-354-774-79	Sequence 79, Appl
39	31	6.6	473	15	US-10-271-012-79	Sequence 79, Appl
40	31	6.6	473	16	US-10-729-122-79	Sequence 79, Appl
41	31	6.6	473	16	US-10-729-039-79	Sequence 79, Appl
42	26	5.5	448	12	US-10-272-898-73	Sequence 73, Appl
43	26	5.5	448	15	US-10-354-774-73	Sequence 73, Appl
44	26	5.5	448	15	US-10-271-012-73	Sequence 73, Appl
45	26	5.5	448	16	US-10-729-122-73	Sequence 73, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-46
; Sequence 46, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-272-898-46

  Query Match            100.0%;  Score 472;  DB 12;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 2
US-10-354-774-46
; Sequence 46, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-354-774-46

  Query Match            100.0%;  Score 472;  DB 15;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 3
US-10-271-012-46
; Sequence 46, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-271-012-46

  Query Match             100.0%;  Score 472;  DB 15;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 4
US-10-729-122-46
; Sequence 46, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-729-122-46

  Query Match             100.0%;  Score 472;  DB 16;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 5
US-10-729-039-46
```

```
; Sequence 46, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-729-039-46

  Query Match            100.0%;  Score 472;  DB 16;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 6
US-10-360-101-220
; Sequence 220, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
;  APPLICANT: Moll, Gert  N.
;  APPLICANT:  Leenhouts, Cornelis J.
;  TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
;  FILE REFERENCE: 2183-5673
;  CURRENT APPLICATION NUMBER: US/10/360,101
;  CURRENT FILING DATE:  2003-02-07
;  PRIOR APPLICATION NUMBER: EP 02077060.8
;  PRIOR FILING DATE: 2002-05-24
;  NUMBER OF SEQ ID NOS: 309
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 220
;   LENGTH: 1290
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: sequence of clostridium botulinum toxin type B
US-10-360-101-220

  Query Match            94.5%;  Score 446;  DB 15;  Length 1290;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 7
US-10-452-024-119
; Sequence 119, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
```

```
ID   AAW68398 standard; protein; 451 AA.
XX
AC   AAW68398;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type D toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW   botulism; BotD.
XX
OS   Clostridium botulinum; serotype D.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30591.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 47; Page 352-353; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC   (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC   BotD soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 15
AAW68396
ID   AAW68396 standard; protein; 452 AA.
XX
AC   AAW68396;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain NCTC 11219.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


Search completed: September 10, 2004, 05:05:55
Job time : 28.6553 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 7.73869 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	324	68.6	1169	4	US-09-255-829-20	Sequence 20, Appl
2	26	5.5	462	1	US-08-480-604A-26	Sequence 26, Appl
3	26	5.5	462	2	US-08-405-496A-26	Sequence 26, Appl
4	26	5.5	462	3	US-08-915-136-26	Sequence 26, Appl
5	26	5.5	462	4	US-09-084-517-26	Sequence 26, Appl
6	23	4.9	23	1	US-08-480-604A-24	Sequence 24, Appl
7	23	4.9	23	2	US-08-405-496A-24	Sequence 24, Appl
8	23	4.9	23	3	US-08-915-136-24	Sequence 24, Appl
9	23	4.9	23	4	US-09-084-517-24	Sequence 24, Appl
10	22	4.7	22	4	US-09-660-742-3	Sequence 3, Appli
11	22	4.7	53	2	US-08-651-818A-19	Sequence 19, Appl
12	22	4.7	53	3	US-09-184-826-19	Sequence 19, Appl
13	22	4.7	54	2	US-08-651-818A-23	Sequence 23, Appl
14	22	4.7	54	3	US-09-184-826-23	Sequence 23, Appl
15	22	4.7	323	1	US-08-591-989-5	Sequence 5, Appli
16	21	4.4	21	2	US-08-651-818A-21	Sequence 21, Appl
17	21	4.4	21	3	US-09-184-826-21	Sequence 21, Appl
18	17	3.6	24	3	US-08-584-031-8	Sequence 8, Appli
19	17	3.6	24	3	US-08-780-496-8	Sequence 8, Appli
20	17	3.6	24	4	US-08-928-069-9	Sequence 9, Appli
21	15	3.2	20	4	US-09-186-489-6	Sequence 6, Appli
22	15	3.2	20	4	US-10-043-665B-6	Sequence 6, Appli
23	13	2.8	858	4	US-09-255-829-22	Sequence 22, Appl
24	13	2.8	858	4	US-09-255-829-29	Sequence 29, Appl
25	12	2.5	16	5	PCT-US94-04361-56	Sequence 56, Appl
26	12	2.5	17	2	US-08-600-783-13	Sequence 13, Appl
27	11	2.3	20	4	US-09-674-677-34	Sequence 34, Appl
28	11	2.3	24	4	US-09-292-858B-27	Sequence 27, Appl
29	11	2.3	254	3	US-09-167-434-7	Sequence 7, Appli
30	11	2.3	254	3	US-08-853-755-7	Sequence 7, Appli
31	11	2.3	302	2	US-08-203-532F-4	Sequence 4, Appli
32	11	2.3	302	3	US-08-950-860-16	Sequence 16, Appl
33	11	2.3	302	3	US-09-078-465-4	Sequence 4, Appli
34	11	2.3	302	5	PCT-US95-01882A-4	Sequence 4, Appli
35	11	2.3	303	2	US-08-203-532F-2	Sequence 2, Appli
36	11	2.3	303	3	US-09-078-465-2	Sequence 2, Appli
37	11	2.3	303	5	PCT-US95-01882A-2	Sequence 2, Appli
38	11	2.3	363	4	US-09-328-352-4930	Sequence 4930, Ap
39	11	2.3	414	5	PCT-US92-06840-2	Sequence 2, Appli
40	10	2.1	25	4	US-09-721-154-14	Sequence 14, Appl
41	10	2.1	335	3	US-09-095-117-2	Sequence 2, Appli
42	10	2.1	335	4	US-09-850-260-2	Sequence 2, Appli
43	10	2.1	449	2	US-08-927-394-2	Sequence 2, Appli
44	10	2.1	480	1	US-07-882-292-2	Sequence 2, Appli
45	10	2.1	480	2	US-08-331-644-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 68.6%; Score 324; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 1.7e-304;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy          87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy         147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy         207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy         267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy         327 IIRRKSNSQSINDDIVRKEDYIYL 350
               ||||||||||||||||||||||||
Db        1146 IIRRKSNSQSINDDIVRKEDYIYL 1169
```

```
RESULT 2
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             5.5%;  Score 26;  DB 1;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 7e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 3
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-08-405-496A-26

Query Match 5.5%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 7e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 4
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OE INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-26

Query Match 5.5%; Score 26; DB 3; Length 462;
Best Local Similarity 100.0%; Pred. No. 7e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 5
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-26

Query Match 5.5%; Score 26; DB 4; Length 462;
Best Local Similarity 100.0%; Pred. No. 7e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

```
RESULT 6
US-08-480-604A-24
; Sequence 24, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-24

  Query Match             4.9%;  Score 23;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23
```

```
RESULT 7
US-08-405-496A-24
; Sequence 24, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-24

  Query Match             4.9%;  Score 23;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23
```

```
RESULT 8
US-08-915-136-24
; Sequence 24, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
```

This Page Blank (uspto)

```
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-23

  Query Match             4.7%;  Score 22;  DB 2;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 14
US-09-184-826-23
; Sequence 23, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-23

  Query Match             4.7%;  Score 22;  DB 3;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 15
US-08-591-989-5
; Sequence 5, Application US/08591989
; Patent No. 5795721
;  GENERAL INFORMATION:
;    APPLICANT:  Ross S. Rabin, Sumedha Jayasena
;    APPLICANT:  and Larry Gold
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC
;    TITLE OF INVENTION:  ACID LIGANDS OF ICP4
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 East Prentice Avenue, Suite #200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.40 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX 49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-591-989-5

  Query Match             4.7%;  Score 22;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db         17 MGHHHHHHHHHHSSGHIEGRHM 38


Search completed: September 10, 2004, 05:17:05
Job time : 8.73869 secs
```

```
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Polypeptide
US-09-660-742-3

  Query Match             4.7%;  Score 22;  DB 4;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 11
US-08-651-818A-19
; Sequence 19, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-19

  Query Match             4.7%;  Score 22;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 12
US-09-184-826-19
; Sequence 19, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-19

  Query Match             4.7%;  Score 22;  DB 3;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 13
US-08-651-818A-23
; Sequence 23, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
```

```
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-24

  Query Match             4.9%;  Score 23;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 9
US-09-084-517-24
; Sequence 24, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-24

  Query Match             4.9%;  Score 23;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 10
US-09-660-742-3
; Sequence 3, Application US/09660742
; Patent No. 6485977
; GENERAL INFORMATION:
;  APPLICANT:  Collmer, Alan
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  RECOMBINANT CONSTRUCTS AND TECHNIQUES FOR DELIVERING TO
;  TITLE OF INVENTION:  EUCARYOTIC CELLS BACTERIAL PROTEINS THAT ARE SECRETED
;  TITLE OF INVENTION:  VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE:  19603/4021
;  CURRENT APPLICATION NUMBER: US/09/660,742
;  CURRENT FILING DATE:  2000-09-13
;  PRIOR APPLICATION NUMBER:  60/153,507
;  PRIOR FILING DATE:  1999-09-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 22
```

```
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 119
;   LENGTH: 1290
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-119

  Query Match             94.5%;  Score 446;  DB 15;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 8
US-10-272-898-42
; Sequence 42, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-272-898-42

  Query Match             94.5%;  Score 446;  DB 12;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 9
US-10-354-774-42
; Sequence 42, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
```

```
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-354-774-42

  Query Match             94.5%;  Score 446;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 10
US-10-271-012-42

; Sequence 42, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-271-012-42

  Query Match             94.5%;  Score 446;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 11
US-10-452-024-2
; Sequence 2, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-2

  Query Match             94.5%;  Score 446;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 12
US-10-452-024-118
; Sequence 118, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 118
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-118

  Query Match             94.5%;  Score 446;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 13
US-10-729-122-42
; Sequence 42, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,122
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1291 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-729-122-42

  Query Match             94.5%;  Score 446;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 14
US-10-729-039-42
; Sequence 42, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1291 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-729-039-42

  Query Match             94.5%;  Score 446;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 15
US-09-910-186A-8
; Sequence 8, Application US/09910186A
```

; Publication No. US20030009025A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Material Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; FILE REFERENCE: A33626-A 067252.0107
; CURRENT APPLICATION NUMBER: US/09/910,186A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; OTHER INFORMATION: sequence
US-09-910-186A-8

Query Match 74.2%; Score 350; DB 10; Length 440;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          35 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQ 94
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           3 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQ 62

Qy          95 NQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW 154
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          63 NQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIW 122

Qy         155 TLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIRE 214
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         123 TLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIRE 182

Qy         215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         183 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 242

Qy         275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         243 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 302

Qy         335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDE 384
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         303 QSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDE 352
```

Search completed: September 10, 2004, 05:52:35
Job time : 25.876 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 6.53889 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	446	94.5	1291	1	A48940	bontoxilysin (EC 3
2	70	14.8	1291	2	I40631	non-proteolytic bo
3	11	2.3	302	2	A55641	homeotic protein G
4	11	2.3	303	2	A56837	homeotic protein M
5	11	2.3	303	2	B49122	homeobox protein M
6	11	2.3	303	2	A48130	growth arrest-spec
7	11	2.3	354	2	S39406	homeotic protein o
8	11	2.3	355	2	S35345	otx1 protein - mou
9	11	2.3	355	2	I56547	homeodomain protei
10	11	2.3	414	2	A40350	transcription repr
11	11	2.3	414	2	A48273	delta/YY1/NF-E1/UC
12	11	2.3	420	2	T39712	hypothetical prote
13	11	2.3	427	2	T42516	hypothetical prote
14	11	2.3	1180	2	S69205	stripe a/b protein
15	10	2.1	132	2	I39193	gene HOXA1 protein
16	10	2.1	133	2	B30242	stem cell protein
17	10	2.1	222	2	I39192	gene HOXA1 protein
18	10	2.1	236	2	S41512	Brn-3b protein - m
19	10	2.1	298	2	S41469	homeotic protein M
20	10	2.1	309	2	T29293	hypothetical prote
21	10	2.1	321	2	T02987	myb-related protei
22	10	2.1	326	2	D83483	probable metal tra
23	10	2.1	331	2	A30242	homeotic protein E
24	10	2.1	335	2	G01448	homeobox protein H
25	10	2.1	351	1	KGZQHL	histidine-rich gly
26	10	2.1	375	2	A46390	cAMP receptor subt
27	10	2.1	387	2	A47446	HNF-3/fork head fa
28	10	2.1	390	1	A38565	polycomb (Pc) prot
29	10	2.1	391	2	H86187	hypothetical prote
30	10	2.1	407	2	G84783	probable pectinest
31	10	2.1	410	2	I38502	gene Brn-3b protei
32	10	2.1	411	2	I58156	Brn-3.2 - mouse
33	10	2.1	430	2	S66671	neuron-derived rec
34	10	2.1	451	2	A55909	transforming prote
35	10	2.1	469	2	I37451	HBF-G2 (HFK-2) pro
36	10	2.1	476	2	A54743	transcription fact
37	10	2.1	480	2	JH0672	brain factor 1 pro
38	10	2.1	580	2	T46024	hypothetical prote
39	10	2.1	623	1	S33167	gene pointed prote
40	10	2.1	625	2	S71930	neuron-derived rec
41	10	2.1	628	2	JC2493	neuron derived orp
42	10	2.1	633	1	A26030	serine/threonine-s
43	10	2.1	657	2	A29454	knob-associated hi
44	10	2.1	665	2	S02358	lamin A - African
45	10	2.1	717	2	S38177	SSV7 protein homol

ALIGNMENTS

RESULT 1
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publicat
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisior
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus tox
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein

A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 94.5%; Score 446; DB 1; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 2
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic si
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 14.8%; Score 70; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.2e-61;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        183 WFFVTITNNL 192
              ||||||||||
Db       1002 WFFVTITNNL 1011
```

RESULT 3
A55641
homeotic protein GAX - human
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 24-Sep-1999
C;Accession: A55641
R;LePage, D.F.; Altomare, D.A.; Testa, J.R.; Walsh, K.
Genomics 24, 535-540, 1994
A;Title: Molecular cloning and localization of the human GAX gene to 7p21.
A;Reference number: A55641; MUID:95229154; PMID:7713505
A;Accession: A55641
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <LEP>
A;Cross-references: GB:L36328; NID:g557549; PIDN:AAA58497.1; PID:g557550
C;Genetics:
A;Gene: GAX

```
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;186-242/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 0.0076;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 4
A56837
homeotic protein MOX2 - human
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 24-Sep-1999
C;Accession: A56837; S52168
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.S.; Theodorakis, K.; Mankoo, B.; Pachnis, V.;
Genomics 26, 550-555, 1995
A;Title: Isolation of the human MOX2 homeobox gene and localization to chromosome 7p22.1
A;Reference number: A56837; MUID:95331791; PMID:7607679
A;Accession: A56837
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-303 <GRI>
A;Cross-references: EMBL:X82629; NID:g732790; PIDN:CAA57949.1; PID:g732791
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.; Theodorakis, K.; Maukoo, B.; Pachnis, V.; K
submitted to the EMBL Data Library, November 1994
A;Description: Isolation of the human MOX2 homeobox gene and localization to chromosome
A;Reference number: S52168
A;Accession: S52168
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-7,'R',9-110,'R',112-122,'Q',124-157,'V',159-303 <GR2>
A;Cross-references: EMBL:X82629
C;Genetics:
A;Gene: GDB:MOX2
A;Cross-references: GDB:120703; OMIM:155970
A;Map position: 3q13-3q13
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0077;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
B49122
homeobox protein Mox-2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Dec-1993 #sequence_revision 17-Nov-1995 #text_change 17-Nov-2000
C;Accession: B49122; S41779; S29902
R;Candia, A.F.; Hu, J.; Crosby, J.; Lalley, P.A.; Noden, D.; Nadeau, J.H.; Wright, C.V.
Development 116, 1123-1136, 1992
A;Title: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially e
A;Reference number: A49122; MUID:93201999; PMID:1363541
A;Accession: B49122
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN1>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
A;Experimental source: C57BL
A;Note: the complete translation is not shown
R;Candia, A.F.; Kovalik, J.P.; Wright, C.V.E.
Nucleic Acids Res. 21, 4982, 1993
A;Title: Amino acid sequence of Mox-2 and comparison to its Xenopus and rat homologs.
A;Reference number: S41469; MUID:94232829; PMID:7909944
A;Accession: S41779
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN2>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
R;Candia, A.
submitted to the EMBL Data Library, October 1992
A;Reference number: S29902
A;Accession: S29902
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 186-246 <CAN>
A;Cross-references: EMBL:Z16406
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0077;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 6
A48130
growth arrest-specific homeobox protein Cvx - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A48130; S31976
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
Mol. Cell. Biol. 13, 3722-3733, 1993
A;Title: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated du
A;Reference number: A48130; MUID:93268321; PMID:8098844
A;Accession: A48130
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-303 <GOR>
A;Cross-references: GB:Z17223; NID:g57951; PIDN:CAA78931.1; PID:g57952
A;Experimental source: aorta
A;Note: sequence extracted from NCBI backbone (NCBIN:132842, NCBIP:132843)
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
submitted to the EMBL Data Library, October 1992
A;Description: Cvx: A diverged homeodomain gene which is rapidly down-regulated followin
A;Reference number: S31976
A;Accession: S31976
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-182,'S',184-303 <GO2>
A;Cross-references: EMBL:Z17223
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0077;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 7
S39406
homeotic protein otx1 - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 15-Oct-1999
C;Accession: S39406
```

```
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; E
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S39406
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-354 <SIM>
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.0088;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
S35345
otx1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000
C;Accession: S35345; S29087
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; E
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S35345
A;Molecule type: DNA
A;Residues: 1-355 <SIM>
R;Simeone, A.; Acampora, D.; Gulisano, M.; Stornaiuolo, A.; Boncinelli, E.
Nature 358, 687-690, 1992
A;Title: Nested expression domains of four homeobox genes in developing rostral brain.
A;Reference number: S29084; MUID:92357147; PMID:1353865
A;Accession: S29087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 31-98 <SI2>
A;Cross-references: EMBL:X68883; NID:g53540; PIDN:CAA48754.1; PID:g53541
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 9
I56547
homeodomain protein otx1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C;Accession: I56547
R;Frantz, G.D.; Weimann, J.M.; Levin, M.E.; McConnell, S.K.
J. Neurosci. 14, 5725-5740, 1994
A;Title: Otx1 and Otx2 define layers and regions in developing cerebral cortex and cereb
A;Reference number: I56547; MUID:95016961; PMID:7931541
A;Accession: I56547
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: GB:L32602; NID:g535739; PIDN:AAA53557.1; PID:g535740
C;Genetics:
A;Gene: OTX1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

  Query Match             2.3%;  Score 11;  DB 2;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 10
A40350
transcription repressor protein YY1 - human
N;Alternate names: transcription repressor/activator protein NF-E1
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Oct-1999
C;Accession: A40350; S78494; S33712; A56419
R;Shi, Y.; Seto, E.; Chang, L.S.; Shenk, T.
Cell 67, 377-388, 1991
A;Title: Transcriptional repression by YY1, a human GLI-Krueppel-related protein, and rel
A;Reference number: A40350; MUID:92005716; PMID:1655281
A;Accession: A40350
A;Molecule type: mRNA
A;Residues: 1-414 <SHI>
A;Cross-references: GB:M77698; NID:g186767; PIDN:AAA59467.1; PID:g186768
A;Experimental source: HeLa cells
A;Note: the authors translated the codon CGC for residue 371 as Lys, CGA for residue 375
R;Whitson, R.H.; Huang, T.; Dang, J.; Itakura, K.
submitted to the EMBL Data Library, July 1992
A;Description: Observed and predicted DNA binding of a zinc finger protein which recogni
A;Reference number: S78494
A;Accession: S78494
A;Molecule type: mRNA
A;Residues: 1-195,'G',197-414 <WHI>
A;Cross-references: EMBL:Z14077; NID:g38010; PIDN:CAA78455.1; PID:g38011
R;Park, K.; Atchison, M.L.
Proc. Natl. Acad. Sci. U.S.A. 88, 9804-9808, 1991
A;Title: Isolation of a candidate repressor/activator, NF-E1 (YY-1), delta, that binds to
A;Reference number: A56419; MUID:92052179; PMID:1946405
A;Accession: S33712
A;Molecule type: mRNA
A;Residues: 1-64,'R',66-195,'G',197-414 <PAR>
A;Cross-references: GB:M76541; NID:g189173; PIDN:AAA59926.1; PID:g189174
C;Keywords: DNA binding; transcription regulation; zinc finger
F;298-320/Region: zinc finger CCHH motif
F;327-347/Region: zinc finger CCHH motif
F;355-377/Region: zinc finger CCHH motif
F;385-407/Region: zinc finger CCHH motif

  Query Match             2.3%;  Score 11;  DB 2;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 11
A48273
delta/YY1/NF-E1/UCRBP transcription factor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 01-Dec-2000
C;Accession: A48273; A42055; A56418
R;Safrany, G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 5559-5563, 1993
```

```
A;Title: Characterization of the mouse gene that encodes the delta/YY1/NF-E1/UCRBP trans
A;Reference number: A48273; MUID:93296177; PMID:8516301
A;Accession: A48273
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-414 <RES>
A;Cross-references: GB:L13968; NID:g293847; PIDN:AAA40477.1; PID:g293849
R;Flanagan, J.R.; Becker, K.G.; Ennist, D.L.; Gleason, S.L.; Driggers, P.H.; Levi, B.Z.;
Mol. Cell. Biol. 12, 38-44, 1992
A;Title: Cloning of a negative transcription factor that binds to the upstream conserved
A;Reference number: A42055; MUID:92107191; PMID:1309593
A;Accession: A42055
A;Molecule type: mRNA
A;Residues: 1-414 <FLA>
A;Cross-references: GB:M73963; NID:g202270; PIDN:AAA40522.1; PID:g202271
A;Note: sequence extracted from NCBI backbone (NCBIN:74641, NCBIP:74642)
R;Hariharan, N.; Kelley, D.E.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 88, 9799-9803, 1991
A;Title: delta, a transcription factor that binds to downstream elements in several poly
A;Reference number: A56418; MUID:92052178; PMID:1946404
A;Accession: A56418
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218,'S',220-374,'G',376-414 <HAR>
A;Cross-references: GB:M74590; NID:g192940; PIDN:AAA37521.1; PID:g192941
C;Genetics:
A;Introns: 227/1; 281/2; 301/3; 354/3
C;Keywords: transcription factor; zinc finger

  Query Match             2.3%;  Score 11;  DB 2;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 12
T39712
hypothetical protein SPBC17D11.01 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: hypothetical protein SPBC20F10.11
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Jan-2000
C;Accession: T39712; T39882
R;Beck, A.; Reinhardt, R.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21872
A;Accession: T39712
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-420 <BEC>
A;Cross-references: EMBL:AL031739; PIDN:CAA21072.1; GSPDB:GN00067; SPDB:SPBC17D11.01
A;Experimental source: strain 972h-; cosmid c17D11
R;Lyne, M.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21888
A;Accession: T39882
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-281 <LYN>
A;Cross-references: EMBL:AL021747; PIDN:CAA16851.1; GSPDB:GN00067
C;Genetics:
A;Gene: SPBC20F10.11; SPDB:SPBC17D11.01
A;Map position: 2
A;Introns: 10/1; 28/2
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match             2.3%;  Score 11;  DB 2;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 13
T42516
hypothetical protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T42516
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42516
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-427 <YOS>
A;Cross-references: EMBL:D89156; NID:g1749519; PIDN:BAA13818.1; PID:g1749520
A;Experimental source: strain PR745
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match             2.3%;  Score 11;  DB 2;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 14
S69205
stripe a/b protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S69205
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube gui
A;Reference number: S69205; MUID:96203082; PMID:8612588
A;Accession: S69205
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1180 <FRO>
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789
C;Keywords: alternative splicing

  Query Match             2.3%;  Score 11;  DB 2;  Length 1180;
  Best Local Similarity   100.0%;  Pred. No. 0.027;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db        641 HHHHHHHHHHS 651


RESULT 15
I39193
gene HOXA1 protein - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 15-Oct-1999
C;Accession: I39193
R;Chariot, A.; Moreau, L.; Senterre, G.; Sobel, M.; Castronovo, V.
Biochem. Biophys. Res. Commun. 215, 713-720, 1995
A;Title: Retinoic acid induces three newly cloned HOXA1 transcripts in MCF7 breast cancer
A;Reference number: I39192; MUID:96011836; PMID:7488013
A;Accession: I39193
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-132 <RES>
A;Cross-references: EMBL:U37431; NID:g1051226; PIDN:AAC50250.1; PID:g1051229
C;Genetics:
```

A;Gene: GDB:HOXA1
A;Cross-references: GDB:120652; OMIM:142955
A;Map position: 7p15.3-7p15.3
A;Introns: 113/3
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 2.1%; Score 10; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 HHHHHHHHHH 12
              ||||||||||
Db         60 HHHHHHHHHH 69
```

Search completed: September 10, 2004, 05:14:45
Job time : 6.63889 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 4.07931 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	446	94.5	1290	1	BXB_CLOBO	P10844 clostridium
2	11	2.3	303	1	MOX2_HUMAN	P50222 homo sapien
3	11	2.3	303	1	MOX2_MOUSE	P32443 mus musculu
4	11	2.3	303	1	MOX2_RAT	P39020 rattus norv
5	11	2.3	354	1	OTX1_HUMAN	P32242 homo sapien
6	11	2.3	355	1	OTX1_MOUSE	P80205 mus musculu
7	11	2.3	355	1	OTX1_RAT	Q63410 rattus norv
8	11	2.3	414	1	TYY1_HUMAN	P25490 homo sapien
9	11	2.3	414	1	TYY1_MOUSE	Q00899 mus musculu
10	11	2.3	420	1	YBE1_SCHPO	O42980 schizosacch
11	11	2.3	466	1	ZIC3_MOUSE	Q62521 mus musculu
12	11	2.3	467	1	ZIC3_HUMAN	O60481 homo sapien
13	11	2.3	611	1	SNF1_CANGA	Q00372 candida gla
14	11	2.3	619	1	SNF1_CANTR	O94168 candida tro
15	10	2.1	202	1	HAN1_CHICK	Q90691 gallus gall
16	10	2.1	204	1	HAN1_SHEEP	Q28555 ovis aries
17	10	2.1	298	1	MOX2_XENLA	P39021 xenopus lae
18	10	2.1	331	1	HXA1_MOUSE	P09022 mus musculu
19	10	2.1	335	1	HXA1_HUMAN	P49639 homo sapien
20	10	2.1	351	1	HRPX_PLALO	P04929 plasmodium
21	10	2.1	375	1	CAR2_DICDI	P34907 dictyosteli
22	10	2.1	387	1	QIN_AVIS3	P56260 avian sarco
23	10	2.1	390	1	PC_DROME	P26017 drosophila
24	10	2.1	410	1	PO42_HUMAN	Q12837 homo sapien
25	10	2.1	411	1	PO42_MOUSE	Q63934 mus musculu
26	10	2.1	430	1	NOR2_RAT	Q63516 rattus norv
27	10	2.1	449	1	GAT6_HUMAN	Q92908 homo sapien
28	10	2.1	451	1	FXGB_CHICK	Q90964 gallus gall
29	10	2.1	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
30	10	2.1	465	1	HNF6_MOUSE	O08755 mus musculu
31	10	2.1	465	1	HNF6_RAT	P70512 rattus norv
32	10	2.1	469	1	FXGA_HUMAN	P55316 homo sapien
33	10	2.1	477	1	FXGB_HUMAN	P55315 homo sapien
34	10	2.1	480	1	FXGB_RAT	Q00939 rattus norv
35	10	2.1	481	1	FXGB_MOUSE	Q60987 mus musculu
36	10	2.1	551	1	CBX4_MOUSE	O55187 mus musculu
37	10	2.1	558	1	CBX4_HUMAN	O00257 homo sapien
38	10	2.1	620	1	SNF1_CANAL	P52497 candida alb
39	10	2.1	623	1	PNT1_DROME	P51022 drosophila
40	10	2.1	626	1	NR43_HUMAN	Q92570 homo sapien
41	10	2.1	628	1	NR43_RAT	P51179 rattus norv
42	10	2.1	633	1	SNF1_YEAST	P06782 saccharomyc
43	10	2.1	657	1	KNOB_PLAFN	P06719 plasmodium
44	10	2.1	665	1	LAMA_XENLA	P11048 xenopus lae
45	10	2.1	717	1	UBPB_YEAST	P36026 saccharomyc

ALIGNMENTS

RESULT 1
BXB_CLOBO
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence.";
RL Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.";
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;

```
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             94.5%;  Score 446;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 2
MOX2_HUMAN
ID   MOX2_HUMAN     STANDARD;      PRT;   303 AA.
AC   P50222; Q9UPL6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE   homeobox).
GN   MEOX2 OR MOX2 OR GAX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95331791; PubMed=7607679;
RA   Grigoriou M., Kastrinaki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA   Pachnis V., Karagogeos D.;
RT   "Isolation of the human MOX2 homeobox gene and localization to
RT   chromosome 7p22.1-p21.3.";
RL   Genomics 26:550-555(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95229154; PubMed=7713505;
RA   Lepage D.F., Walsh K.;
RT   "Molecular cloning and localization of the human GAX gene to 7p21.";
RL   Genomics 24:535-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
```

```
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 230-303 FROM N.A.
RA   Cordes M., Lacy M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Embryo and placenta.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82629; CAA57949.1; -.
DR   EMBL; L36328; AAA58497.1; -.
DR   EMBL; BC017021; AAH17021.1; -.
DR   EMBL; AC004452; AAC06184.1; -.
DR   PIR; A55641; A55641.
DR   PIR; A56837; A56837.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04005; -.
DR   Genew; HGNC:7014; MEOX2.
DR   MIM; 600535; -.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DNA_BIND    186    245       HOMEOBOX.
FT   CONFLICT     58     58       G -> D (IN REF. 2).
FT   CONFLICT     79     79       MISSING (IN REF. 2).
SQ   SEQUENCE   303 AA;  33457 MW;  809ADE0CD090023D CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0097;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 3
MOX2_MOUSE
ID   MOX2_MOUSE     STANDARD;      PRT;   303 AA.
AC   P32443;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201999; PubMed=1363541;
RA   Candia A.F., Hu J., Crosby J., Lalley P.A., Noden D., Nadeau J.H.,
RA   Wright C.V.E.;
RT   "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are
RT   differentially expressed during early mesodermal patterning in mouse
RT   embryos.";
RL   Development 116:1123-1136(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94232829; PubMed=7909944;
RA   Candia A.F., Kovalik J.-P., Wright C.V.E.;
RT   "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat
RT   homologs.";
RL   Nucleic Acids Res. 21:4982-4982(1993).
RN   [3]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=95349593; PubMed=7623821;
RA   Andres V., Fisher S., Wearsch P., Walsh K.;
RT   "Regulation of Gax homeobox gene transcription by a combination of
RT   positive factors including myocyte-specific enhancer factor 2.";
RL   Mol. Cell. Biol. 15:4272-4281(1995).
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-
CC       8.5 dpc it is found on the entire epithelium of the somite. At 9.5
CC       dpc its expression is restricted to the sclerotome. At 10.5 dpc it
CC       is found in sclerotomally derived cells including the vertebral
CC       and costal precursors.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16406; CAA78899.1; -.
DR   EMBL; S79168; -; NOT_ANNOTATED_CDS.
DR   PIR; B49122; B49122.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04048; -.
DR   MGD; MGI:103219; Meox2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
```

```
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       63     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33506 MW;  41BD05FC39AA4427 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0097;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 4
MOX2_RAT
ID   MOX2_RAT       STANDARD;      PRT;   303 AA.
AC   P39020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Growth arrest-specific homeobox).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93268321; PubMed=8098844;
RA   Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,
RA   Walsh K.;
RT   "Molecular cloning of a diverged homeobox gene that is rapidly down-
RT   regulated during the G0/G1 transition in vascular smooth muscle
RT   cells.";
RL   Mol. Cell. Biol. 13:3722-3733(1993).
RN   [2]
RP   REVISIONS.
RA   Walsh K.;
RL   Submitted (MAR-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Aorta and heart. Also detected in lung and
CC       kidney.
CC   -!- INDUCTION: Rapidly and transiently down-regulated during the
CC       transition from G0 to G1 induced by mitogen stimulation.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z17223; CAA78931.1; -.
DR   PIR; A48130; A48130.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04054; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       64     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33605 MW;  7776642AEFA3A2E8 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0097;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
OTX1_HUMAN
ID   OTX1_HUMAN     STANDARD;      PRT;   354 AA.
AC   P32242;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
```

```
CC       the sense organs. Can bind to the BCD target sequence (BTS):
CC       5'-TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC007621; AAH07621.1; -.
DR   PIR; S39406; S39406.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02079; -.
DR   Genew; HGNC:8521; OTX1.
DR   MIM; 600036; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    301       HIS-RICH.
SQ   SEQUENCE   354 AA;  37327 MW;  E32C1E47464BDDAA CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 6
OTX1_MOUSE
ID   OTX1_MOUSE     STANDARD;      PRT;   355 AA.
AC   P80205;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1 OR OTX-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 31-98 FROM N.A.
RX   MEDLINE=92357147; PubMed=1353865;
RA   Simeone A., Acampora D., Gulisano M., Stornaiuolo A., Boncinelli E.;
RT   "Nested expression domains of four homeobox genes in developing
RT   rostral brain.";
RL   Nature 358:687-690(1992).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC057105; AAH57105.1; -.
DR   EMBL; BC058354; AAH58354.1; -.
DR   EMBL; X68883; CAA48754.1; -.
DR   PIR; S35345; S35345.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02080; -.
DR   MGD; MGI:97450; Otx1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
```

```
SQ   SEQUENCE   355 AA;  37531 MW;  DA5136D9D4FC948F CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 7
OTX1_RAT
ID   OTX1_RAT       STANDARD;      PRT;   355 AA.
AC   Q63410; Q64203;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95016961; PubMed=7931541;
RA   Frantz G.D., Weimann J.M., Levin M.E., McConnell S.K.;
RT   "Otx1 and Otx2 define layers and regions in developing cerebral
RT   cortex and cerebellum.";
RL   J. Neurosci. 14:5725-5740(1994).
RN   [2]
RP   SEQUENCE OF 176-323 FROM N.A.
RX   MEDLINE=96108898; PubMed=8613727;
RA   Robel L., Ding M., James A.J., Lin X., Simeone A., Leckman J.F.,
RA   Vaccarino F.M.;
RT   "Fibroblast growth factor 2 increases Otx2 expression in precursor
RT   cells from mammalian telencephalon.";
RL   J. Neurosci. 15:7879-7891(1995).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3' (By similarity). May play a role in the specification
CC       or differentiation of neurons in the deep layers of the cerebral
CC       cortex, and also in cerebellar regionalization during early
CC       development.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: Expressed throughout the forebrain and
CC       midbrain during development, and in addition is also seen in
CC       discrete spatial and temporal domains in the developing cerebral
CC       cortex and cerebellum. Confined to a subpopulation of neurons in
CC       layers 5 and 6 within the adult cerebral cortex and during
CC       development expression is high in the progenitors of these deep-
CC       layer cells. Expressed in the developing cerebellum in spatially
CC       restricted regions of the external granular layer.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32602; AAA53557.1; -.
DR   EMBL; S81924; -; NOT_ANNOTATED_CDS.
DR   PIR; I56547; I56547.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
FT   CONFLICT    219    219       T -> A (IN REF. 2).
FT   CONFLICT    229    229       R -> G (IN REF. 2).
SQ   SEQUENCE   355 AA;  37602 MW;  C875871723D0B876 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
TYY1_HUMAN
ID   TYY1_HUMAN     STANDARD;      PRT;   414 AA.
AC   P25490; Q14935;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1).
GN   YY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92005716; PubMed=1655281;
RA   Shi Y., Seto E., Chang L.-S., Shenk T.;
RT   "Transcriptional repression by YY1, a human GLI-Kruppel-related
RT   protein, and relief of repression by adenovirus E1A protein.";
RL   Cell 67:377-388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=92052179; PubMed=1946405;
RA   Park K., Atchison M.;
RT   "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
RT   that binds to the immunoglobulin kappa 3' enhancer and the
RT   immunoglobulin heavy-chain mu E1 site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Whitson R.H., Huang T., Dang J., Itakura K.;
RL   Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
```

```
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98152931; PubMed=9493912;
RA   McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
RA   Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
RT   "Targeting of the YY1 transcription factor to the nucleolus and the
RT   nuclear matrix in situ: the C-terminus is a principal determinant for
RT   nuclear trafficking.";
RL   J. Cell. Biochem. 68:500-510(1998).
RN   [6]
RP   INTERACTION WITH YAF2.
RX   MEDLINE=97169297; PubMed=9016636;
RA   Kalenik J.L., Chen D., Bradley M.E., Chen S.-J., Lee T.-C.;
RT   "Yeast two-hybrid cloning of a novel zinc finger protein that
RT   interacts with the multifunctional transcription factor YY1.";
RL   Nucleic Acids Res. 25:843-849(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
RX   MEDLINE=97098436; PubMed=8942976;
RA   Houbaviy H.B., Usheva A., Shenk T., Burley S.K.;
RT   "Cocrystal structure of YY1 bound to the adeno-associated virus P5
RT   initiator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 353-379.
RX   MEDLINE=98308000; PubMed=9642075;
RA   Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
RA   Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
RT   "Design, synthesis and structure of a zinc finger with an artificial
RT   beta-turn.";
RL   J. Mol. Biol. 279:973-986(1998).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
CC       EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
CC       PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77698; AAA59467.1; -.
DR   EMBL; M76541; AAA59926.1; -.
DR   EMBL; Z14077; CAA78455.1; -.
DR   EMBL; BC037308; AAH37308.1; -.
DR   PIR; A40350; A40350.
DR   PDB; 1UBD; 23-DEC-96.
DR   PDB; 1ZNM; 01-APR-98.
DR   TRANSFAC; T00915; -.
DR   Genew; HGNC:12856; YY1.
DR   MIM; 600013; -.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     69       GLY-RICH.
FT   DOMAIN       70     80       POLY-HIS.
FT   DOMAIN      159    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE 1.
FT   ZN_FING     325    347       C2H2-TYPE 2.
FT   ZN_FING     353    377       C2H2-TYPE 3.
FT   ZN_FING     383    407       C2H2-TYPE 4.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION.
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN.
FT   CONFLICT     65     65       H -> R (IN REF. 2).
FT   CONFLICT    196    196       G -> R (IN REF. 1).
FT   STRAND      296    297
FT   TURN        301    302
FT   STRAND      306    307
FT   HELIX       310    317
FT   HELIX       318    320
FT   STRAND      325    326
FT   TURN        328    330
FT   STRAND      333    334
FT   HELIX       337    343
FT   HELIX       344    346
FT   TURN        347    347
FT   STRAND      353    354
FT   TURN        358    359
FT   STRAND      363    364
FT   HELIX       367    378
FT   TURN        388    389
FT   TURN        397    400
FT   HELIX       401    407
SQ   SEQUENCE   414 AA;  44712 MW;  058C05A0AD2D04E6 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 9
TYY1_MOUSE
ID   TYY1_MOUSE     STANDARD;      PRT;   414 AA.
AC   Q00899;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1) (UCR-motif DNA-binding protein).
GN   YY1 OR UCRBP.
```

```
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107191; PubMed=1309593;
RA   Flanagan J.R., Becker K.G., Ennist D.L., Gleason S.L., Driggers P.H.,
RA   Levi B.-Z., Appella E., Ozato K.;
RT   "Cloning of a negative transcription factor that binds to the
RT   upstream conserved region of Moloney murine leukemia virus.";
RL   Mol. Cell. Biol. 12:38-44(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93296177; PubMed=8516301;
RA   Safrany G., Perry R.P.;
RT   "Characterization of the mouse gene that encodes the delta/YY1/NF-
RT   E1/UCRBP transcription factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5559-5563(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052178; PubMed=1946404;
RA   Hariharan N., Kelley D.E., Perry R.P.;
RT   "Delta, a transcription factor that binds to downstream elements in
RT   several polymerase II promoters, is a functionally versatile zinc
RT   finger protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9799-9803(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. BINDS TO
CC       THE UPSTREAM CONSERVED REGION (UCR) (5'-CGCCATTTT-3') OF MOLONEY
CC       MURINE LEUKEMIA VIRUS (MULV).
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73963; AAA40522.1; -.
DR   EMBL; L13968; AAA40477.1; -.
DR   EMBL; L13969; AAA40477.1; JOINED.
DR   EMBL; L13965; AAA40477.1; JOINED.
DR   EMBL; L13966; AAA40477.1; JOINED.
DR   EMBL; L13967; AAA40477.1; JOINED.
DR   EMBL; M74590; AAA37521.1; -.
DR   EMBL; BC055899; AAH55899.1; -.
DR   PIR; A48273; A48273.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T00278; -.
DR   TRANSFAC; T00865; -.
DR   MGD; MGI:99150; Yy1.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003700; F:transcription factor activity; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     70       GLY-RICH.
FT   DOMAIN       71     82       POLY-HIS.
FT   DOMAIN      161    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    407       C2H2-TYPE.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION
FT                                (BY SIMILARITY).
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION (BY SIMILARITY).
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN (BY SIMILARITY).
FT   CONFLICT    219    219       F -> S (IN REF. 3).
FT   CONFLICT    375    375       R -> G (IN REF. 3).
SQ   SEQUENCE   414 AA;  44717 MW;  C012378288E984F9 CRC64;

  Query Match          2.3%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 10
YBE1_SCHPO
ID   YBE1_SCHPO     STANDARD;      PRT;   420 AA.
AC   O42980; P78807;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative thiol protease C17D11.01 (EC 3.4.22.-).
GN   SPBC17D11.01 OR SPBC20F10.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
```

```
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- SIMILARITY: Belongs to peptidase family C48.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031739; CAA21072.1; -.
DR   EMBL; AL021747; CAA16851.1; -.
DR   EMBL; D89156; BAA13818.1; -.
DR   GeneDB_SPombe; SPBC17D11.01; -.
DR   InterPro; IPR003653; Peptidase_C48.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Thiol protease.
FT   DOMAIN      299    312       POLY-HIS.
FT   CONFLICT     57     57       L -> W (IN REF. 2).
FT   CONFLICT    420    420       F -> FGCYTCVV (IN REF. 2).
SQ   SEQUENCE   420 AA;  47257 MW;  2125D94D9A2D9E87 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 11
ZIC3_MOUSE
ID   ZIC3_MOUSE     STANDARD;      PRT;   466 AA.
AC   Q62521;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=96132843; PubMed=8557628;
RA   Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA   Chapman V.M., Mikoshiba K.;
RT   "The mouse zic gene family. Homologues of the Drosophila pair-rule
RT   gene odd-paired.";
RL   J. Biol. Chem. 271:1043-1047(1996).
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: CNS. A high level expression is seen in the
CC       cerebellum.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D70849; BAA11116.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T04671; -.
DR   MGD; MGI:106676; Zic3.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     96       POLY-HIS.
FT   ZN_FING     294    321       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     327    351       C2H2-TYPE 2.
FT   ZN_FING     357    381       C2H2-TYPE 3.
FT   ZN_FING     387    409       C2H2-TYPE 4.
SQ   SEQUENCE   466 AA;  50644 MW;  089749DF726F069B CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 12
ZIC3_HUMAN
ID   ZIC3_HUMAN     STANDARD;      PRT;   467 AA.
AC   O60481;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HTX1 MET-323.
RX   MEDLINE=98016414; PubMed=9354794;
RA   Gebbia M., Ferrero G.B., Pilia G., Bassi M.T., Aylsworth A.S.,
RA   Penman-Splitt M., Bird L.M., Bamforth J.S., Burn J., Schlessiner D.,
RA   Nelson D.L., Casey B.;
RT   "X-linked situs abnormalities result from mutations in ZIC3.";
RL   Nat. Genet. 17:305-308(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in ZIC3 are the cause of X-Linked visceral
CC       heterotaxy (HTX1) [MIM:306955]. HTX1 is a disease characterized by
CC       congenital heart disease and alterations of visceral situs,
CC       including asplenia or polysplenia, symmetric liver, intestinal
CC       malrotation, and abnormal lung lobation.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028706; AAC05594.1; -.
DR   EMBL; AL035443; CAB41648.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:12874; ZIC3.
DR   MIM; 300265; -.
DR   MIM; 306955; -.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein; Disease mutation.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     97       POLY-HIS.
FT   ZN_FING     295    322       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     328    352       C2H2-TYPE 2.
FT   ZN_FING     358    382       C2H2-TYPE 3.
FT   ZN_FING     388    410       C2H2-TYPE 4.
FT   VARIANT     323    323       T -> M (in HTX1).
FT                                /FTId=VAR_007753.
SQ   SEQUENCE   467 AA;  50569 MW;  3150CF13C0679568 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 13
SNF1_CANGA
ID   SNF1_CANGA     STANDARD;      PRT;   611 AA.
AC   Q00372;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCCLS84;
RX   MEDLINE=97101049; PubMed=8945576;
RA   Petter R., Kwon-Chung K.J.;
RT   "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT   pathogenic yeast Candida glabrata.";
RL   Infect. Immun. 64:5269-5273(1996).
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78130; AAB48642.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       6     17       POLY-HIS.
FT   DOMAIN      39    290       PROTEIN KINASE.
FT   NP_BIND     45     53       ATP (BY SIMILARITY).
FT   BINDING     68     68       ATP (BY SIMILARITY).
FT   ACT_SITE   161    161       BY SIMILARITY.
FT   MOD_RES    194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  70049 MW;  89E17812A4900CD0 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db          8 HHHHHHHHHHS 18


RESULT 14
SNF1_CANTR
ID   SNF1_CANTR     STANDARD;      PRT;   619 AA.
AC   O94168;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanai T., Ogawa K., Ueda M., Tanaka A.;
RT   "Genetic evaluation of the function of SNF1 in Candida tropicalis.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB024535; BAA75889.1; -.
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       17     29       POLY-HIS.
FT   DOMAIN       52    303       PROTEIN KINASE.
FT   NP_BIND      58     66       ATP (BY SIMILARITY).
FT   BINDING      81     81       ATP (BY SIMILARITY).
FT   ACT_SITE    174    174       BY SIMILARITY.
FT   MOD_RES     207    207       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  70323 MW;  0FCF1FC3DCE706D7 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db         20 HHHHHHHHHHS 30


RESULT 15
HAN1_CHICK
ID   HAN1_CHICK     STANDARD;      PRT;   202 AA.
AC   Q90691;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Heart- and neural crest derivatives-expressed protein 1
DE   (Extraembryonic tissues, heart, autonomic nervous sytem and neural
DE   crest derivatives-expressed protein 1) (eHAND).
GN   HAND1 OR EHAND.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=96123273; PubMed=8533092;
RA   Srivastava D., Cserjesi P., Olson E.N.;
RT   "A subclass of bHLH proteins required for cardiac morphogenesis.";
RL   Science 270:1995-1998(1995).
CC   -!- FUNCTION: Plays an essential role in cardiac morphogenesis.
CC   -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC       bHLH protein.
CC   -!- DEVELOPMENTAL STAGE: At stage 8, detected in the cardiac crescent.
CC       At stage 10, expressed thoughout the cardiac tube and the sinus
CC       venosus. By stage 15, expressed homogenously in the various
CC       regions of the heart, including the atria, future left ventricle,
CC       bulbus cordis and truncus arteriosus, and in the forming branchial
CC       arches. Expression persists through stage 20, but decreases
CC       thereafter.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U40041; AAC59734.1; -.
DR   TRANSFAC; T04371; -.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS50888; HLH; 1.
KW   Developmental protein; Transcription regulation; DNA-binding;
KW   Nuclear protein.
FT   DOMAIN       9     18       POLY-HIS.
FT   DNA_BIND     82    95       BASIC DOMAIN.
FT   DOMAIN       96    139      HELIX-LOOP-HELIX MOTIF.
SQ   SEQUENCE   202 AA;  22679 MW;  1819896705268AA6 CRC64;

  Query Match             2.1%;  Score 10;  DB 1;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 0.062;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHH 12
              ||||||||||
Db          9 HHHHHHHHHH 18


Search completed: September 10, 2004, 05:07:13
Job time : 5.07931 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 19.1967 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 472
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	107	22.7	1291	2	Q93G71	Q93g71 clostridium
2	70	14.8	1291	2	Q08077	Q08077 clostridium
3	68	14.4	1291	2	Q9ZAJ8	Q9zaj8 clostridium
4	68	14.4	1291	2	Q933K0	Q933k0 clostridium
5	53	11.2	361	2	Q45848	Q45848 clostridium
6	53	11.2	361	2	Q45846	Q45846 clostridium
7	43	9.1	441	2	Q9X708	Q9x708 clostridium
8	43	9.1	1291	2	Q8GR96	Q8gr96 clostridium
9	26	5.5	78	2	Q7WUH9	Q7wuh9 clostridium
10	26	5.5	78	2	Q7WRW0	Q7wrw0 clostridium
11	12	2.5	226	11	Q9QZD2	Q9qzd2 rattus norv
12	12	2.5	516	13	O42186	O42186 brachydanio
13	12	2.5	517	13	Q90YL6	Q90yl6 brachydanio
14	12	2.5	569	13	Q802D2	Q802d2 brachydanio
15	12	2.5	679	13	Q803K9	Q803k9 brachydanio
16	12	2.5	706	5	Q8SXG1	Q8sxg1 drosophila
17	12	2.5	861	5	Q9VRT8	Q9vrt8 drosophila
18	11	2.3	163	5	Q7YTA3	Q7yta3 glomeris ma
19	11	2.3	225	13	Q90XZ3	Q90xz3 erpetoichth
20	11	2.3	230	4	O75263	O75263 homo sapien
21	11	2.3	243	4	Q9P2R2	Q9p2r2 homo sapien
22	11	2.3	243	6	Q9N2G2	Q9n2g2 gorilla gor
23	11	2.3	243	6	Q9N2G3	Q9n2g3 pan troglod
24	11	2.3	243	6	Q9N2G1	Q9n2g1 pongo pygma
25	11	2.3	302	13	Q90YH7	Q90yh7 gallus gall
26	11	2.3	303	6	Q95JA6	Q95ja6 sus scrofa
27	11	2.3	303	11	Q99M23	Q99m23 mus musculu
28	11	2.3	355	11	Q91ZJ3	Q91zj3 mus musculu
29	11	2.3	358	13	Q9DG32	Q9dg32 xenopus lae
30	11	2.3	411	11	Q8CHJ4	Q8chj4 rattus norv
31	11	2.3	441	5	Q8T1W3	Q8t1w3 dictyosteli
32	11	2.3	449	10	Q8L4M9	Q8l4m9 oryza sativ
33	11	2.3	456	11	Q8BSB3	Q8bsb3 mus musculu
34	11	2.3	467	13	Q9YGC6	Q9ygc6 xenopus lae
35	11	2.3	511	10	Q8GXZ5	Q8gxz5 arabidopsis
36	11	2.3	906	5	Q8IN91	Q8in91 drosophila
37	11	2.3	1122	5	Q8SXD4	Q8sxd4 drosophila
38	11	2.3	1147	5	Q95PH8	Q95ph8 dictyosteli
39	11	2.3	1180	5	Q24163	Q24163 drosophila
40	11	2.3	1186	5	Q9VED3	Q9ved3 drosophila
41	11	2.3	1283	5	Q95RH4	Q95rh4 drosophila
42	11	2.3	1911	5	Q9W4M7	Q9w4m7 drosophila
43	10	2.1	53	2	O55244	O55244 escherichia
44	10	2.1	233	4	Q8TC53	Q8tc53 homo sapien
45	10	2.1	233	5	Q8MP30	Q8mp30 dictyosteli

ALIGNMENTS

RESULT 1
Q93G71
ID Q93G71 PRELIMINARY; PRT; 1291 AA.
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF295926; AAK97132.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 22.7%; Score 107; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 8.6e-99;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        256 YKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYI 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 YKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYI 1134

Qy        316 NYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVY 362
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 NYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVY 1181
```

RESULT 2
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR PIR; I40631; I40631.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 14.8%; Score 70; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 2.2e-61;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        183 WFFVTITNNL 192
              ||||||||||
Db       1002 WFFVTITNNL 1011
```

RESULT 3
Q9ZAJ8
ID Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 14.4%; Score 68; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 2.3e-59;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        364 YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHR 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHR 1242

Qy        424 FYESGIVF 431
              ||||||||
Db       1243 FYESGIVF 1250
```

RESULT 4
Q933K0
ID Q933K0 PRELIMINARY; PRT; 1291 AA.
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=593, and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.

```
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             14.4%;  Score 68;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-59;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        364 YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHR 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHR 1242

Qy        424 FYESGIVF 431
              ||||||||
Db       1243 FYESGIVF 1250


RESULT 5
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;

  Query Match             11.2%;  Score 53;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-44;
  Matches   53;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIRED 361


RESULT 6
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;

  Query Match             11.2%;  Score 53;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-44;
  Matches   53;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 IQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIRED 361


RESULT 7
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;

  Query Match              9.1%;  Score 43;  DB 2;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-34;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        388 TIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV 430
              |||||||||||||||||||||||||||||||||||||||||||||
Db        365 TIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV 407


RESULT 8
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             9.1%;  Score 43;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-34;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        388 TIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV 430
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1207 TIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV 1249


RESULT 9
Q7WUH9
ID   Q7WUH9      PRELIMINARY;      PRT;    78 AA.
AC   Q7WUH9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CL196;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327855; AAQ16536.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE   78 AA;  9386 MW;  6E6465A3A32CBDBC CRC64;

  Query Match             5.5%;  Score 26;  DB 2;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-18;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        272 PLMYNKEYYMFNAGNKNSYIKLKKDS 297
              |||||||||||||||||||||||||
Db          1 PLMYNKEYYMFNAGNKNSYIKLKKDS 26


RESULT 10
Q7WRW0
ID   Q7WRW0      PRELIMINARY;      PRT;    78 AA.
AC   Q7WRW0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=215, and 216;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327856; AAQ16537.1; -.
DR   EMBL; AY327857; AAQ16538.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      78     78
SQ   SEQUENCE   78 AA;  9370 MW;  7CF465A3A32CBDBC CRC64;

  Query Match             5.5%;  Score 26;  DB 2;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-18;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        272 PLMYNKEYYMFNAGNKNSYIKLKKDS 297
              |||||||||||||||||||||||||
Db          1 PLMYNKEYYMFNAGNKNSYIKLKKDS 26


RESULT 11
Q9QZD2
ID   Q9QZD2      PRELIMINARY;      PRT;   226 AA.
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
SQ   SEQUENCE   226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

  Query Match             2.5%;  Score 12;  DB 11;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
```

```
                    ||||||||||||
Db          17 HHHHHHHHHHSS 28


RESULT 12
O42186
ID   O42186      PRELIMINARY;      PRT;   516 AA.
AC   O42186;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FTZ-F1 homolog.
GN   NR5A2 OR ZFF1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322094; PubMed=9178748;
RA   Liu D., Le Drean Y., Ekker M., Xiong F., Hew C.L.;
RT   "Teleost FTZ-F1 homolog and its splicing variant determine the
RT   expression of the salmon gonadotropin IIbeta subunit gene.";
RL   Mol. Endocrinol. 11:877-890(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AF014926; AAC60274.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04065; -.
DR   ZFIN; ZDB-GENE-990415-79; nr5a2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   516 AA;  58805 MW;  500DA09E116D5FA5 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 0.0042;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 13
Q90YL6
ID   Q90YL6      PRELIMINARY;      PRT;   517 AA.
AC   Q90YL6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nuclear receptor FF1aE.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chai C., Tan J.H., Gao W., Teh H.L., Chan W.K.;
RT   "Full length zebrafish ff1aE cDNA.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AY034061; AAK54449.1; -.
DR   HSSP; P19793; 1LBD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   517 AA;  58761 MW;  CF48D08AB7A75F29 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.0042;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 14
Q802D2
ID   Q802D2      PRELIMINARY;      PRT;   569 AA.
AC   Q802D2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048886; AAH48886.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
KW   Cyclin.
SQ   SEQUENCE   569 AA;  64117 MW;  812A7332C06EF4C2 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 0.0045;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        486 HHHHHHHHHHSS 497
```

```
RESULT 15
Q803K9
ID   Q803K9      PRELIMINARY;      PRT;   679 AA.
AC   Q803K9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T1 (Fragment).
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044435; AAH44435.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
KW   Cyclin.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  74887 MW;  AB5667D59439659F CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 679;
  Best Local Similarity   100.0%;  Pred. No. 0.0053;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        488 HHHHHHHHHHSS 499


Search completed: September 10, 2004, 05:12:46
Job time : 20.1967 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 31.1347 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 2545
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2545	100.0	472	2	AAW68394	Aaw68394 Clostridi
2	2382	93.6	1070	3	AAY93308	Aay93308 A mangane
3	2382	93.6	1095	3	AAY93311	Aay93311 A mangane
4	2375	93.3	1291	2	AAW68392	Aaw68392 Clostridi
5	2340	91.9	440	3	AAY77135	Aay77135 Synthetic
6	2340	91.9	440	4	AAB04091	Aab04091 Botulism
7	2322	91.2	439	4	AAB04085	Aab04085 Botulism
8	2307	90.6	472	2	AAW68393	Aaw68393 Clostridi
9	2298	90.3	848	4	AAB04082	Aab04082 Botulism
10	1703	66.9	1169	2	AAW56017	Aaw56017 Recombina
11	1369.5	53.8	473	2	AAW68400	Aaw68400 Clostridi
12	1210.5	47.6	449	3	AAY77139	Aay77139 Synthetic
13	1210.5	47.6	449	4	AAB04167	Aab04167 Botulism
14	988.5	38.8	462	2	AAR95009	Aar95009 Type A ne
15	988.5	38.8	462	2	AAW68390	Aaw68390 Clostridi
16	887	34.9	448	2	AAW68399	Aaw68399 Clostridi
17	869	34.1	445	2	AAW68391	Aaw68391 Clostridi
18	864.5	34.0	452	2	AAW68396	Aaw68396 Clostridi
19	847	33.3	451	2	AAW68395	Aaw68395 Clostridi
20	838.5	32.9	437	4	AAB04088	Aab04088 Botulism
21	838.5	32.9	438	2	AAR95008	Aar95008 Type A ne
22	838.5	32.9	438	2	AAW68389	Aaw68389 Clostridi
23	838.5	32.9	438	3	AAY77134	Aay77134 Synthetic
24	838.5	32.9	1067	3	AAY93307	Aay93307 A mangane
25	838.5	32.9	1092	3	AAY93310	Aay93310 A mangane
26	838.5	32.9	1296	2	AAR95010	Aar95010 C. botuli
27	835	32.8	435	4	AAB04090	Aab04090 Botulism
28	833.5	32.8	434	4	AAB04089	Aab04089 Botulism
29	824.5	32.4	432	3	AAY77142	Aay77142 Native bo
30	824.5	32.4	837	3	AAY77140	Aay77140 Native bo
31	819.5	32.2	1295	5	AAU99339	Aau99339 Clostridi
32	817.5	32.1	847	4	AAB04081	Aab04081 Botulism
33	786.5	30.9	415	4	AAB04083	Aab04083 Botulism
34	771.5	30.3	451	2	AAW68398	Aaw68398 Clostridi
35	756	29.7	685	4	AAE07893	Aae07893 Modified
36	755	29.7	660	4	AAE07898	Aae07898 Modified
37	753.5	29.6	862	4	AAE07890	Aae07890 Modified
38	753.5	29.6	887	4	AAE07892	Aae07892 Modified
39	753.5	29.6	979	6	AAE35713	Aae35713 BoNT/F-Hc
40	753.5	29.6	1192	6	AAE35711	Aae35711 BoNT/F-Hc
41	753.5	29.6	1192	6	AAE35710	Aae35710 BoNT/F-Hc
42	753	29.6	1092	4	AAE07900	Aae07900 C. botuli
43	752.5	29.6	645	4	AAE07894	Aae07894 Modified
44	752.5	29.6	645	6	AAE35692	Aae35692 DipT HN d
45	752.5	29.6	657	6	AAE35693	Aae35693 BoNT/F-Hc

ALIGNMENTS

RESULT 1
AAW68394
ID AAW68394 standard; protein; 472 AA.
XX
AC AAW68394;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30581.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 303-305; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant

```
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             100.0%;  Score 2545;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-180;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 2
AAY93308
ID   AAY93308 standard; protein; 1070 AA.
XX
AC   AAY93308;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype B.
XX
OS   Synthetic.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 45-47; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC   be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype B. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a domain
CC   that translocates the SOD of the composition into the neuronal cell.
CC   After translocation, the linker is cleaved to release the SOD. The
CC   composition is useful for treating neuronal diseases caused or augmented
CC   by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC   disease, Huntington's disease and motor neurone diseases
XX
SQ   Sequence 1070 AA;

  Query Match             93.6%;  Score 2382;  DB 3;  Length 1070;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-168;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 684

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 744

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 804

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 864

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 924

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 984

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1044

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1045 EVKRKPYNLKLGCNWQFIPKDEGWTE 1070


RESULT 3
AAY93311
ID   AAY93311 standard; protein; 1095 AA.
XX
AC   AAY93311;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
```

XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype B.
XX
OS Synthetic.
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 54-57; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide dismutase
CC (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC serotype B. The specification describes a composition for delivery of SOD
CC to neuronal cells. The composition comprises SOD linked, by a cleavable
CC linker, to a neuronal cell targeting component (NCTC). This component has
CC a domain that binds to a neuronal cell and a domain that translocates the
CC SOD of the composition into the neuronal cell. After translocation, the
CC linker is cleaved to release the SOD. The composition is useful for
CC treating neuronal diseases caused or augmented by oxidative stress, such
CC as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases
XX
SQ Sequence 1095 AA;

Query Match 93.6%; Score 2382; DB 3; Length 1095;
Best Local Similarity 100.0%; Pred. No. 5.7e-168;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        650 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 709

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        710 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 769

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        770 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 829

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        830 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 889

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 949

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        950 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1009

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1010 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1069

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1070 EVKRKPYNLKLGCNWQFIPKDEGWTE 1095
```

RESULT 4
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30579.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 291-296; 428pp; English.
XX
CC This is the amino acid sequence of the type B toxin of Clostridium
CC botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC B toxin has been expressed as a histidine-tagged protein in Escherichia
CC coli. The invention relates to C. botulinum recombinant toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of the recombinant proteins are E. coli, insect cells and
CC yeast cells. The recombinant proteins are used as immunogens for the
CC production of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1291 AA;

Query Match 93.3%; Score 2375; DB 2; Length 1291;
Best Local Similarity 99.8%; Pred. No. 2.3e-167;
Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYHEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

```
RESULT 5
AAY77135
ID   AAY77135 standard; protein; 440 AA.
XX
AC   AAY77135;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype B (BoNTB) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype B; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87213.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 23; Page; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast. Note: This
CC   sequence is not given in the specification, but is decoded from the BoNTB
CC   Hc DNA sequence given on pages 39-40
XX
SQ   Sequence 440 AA;

  Query Match             91.9%;  Score 2340;  DB 3;  Length 440;
  Best Local Similarity   99.5%;  Pred. No. 2.3e-165;
  Matches  438;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         33 MFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 92
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MANKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 60

Qy         93 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy        153 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 180

Qy        213 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 240

Qy        273 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 300

Qy        333 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIK 392
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        301 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDELYNTIQIK 360

Qy        393 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 420

Qy        453 YNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||
Db        421 YNLKLGCNWQFIPKDEGWTE 440
```

```
RESULT 6
AAB04091
ID   AAB04091 standard; protein; 440 AA.
XX
AC   AAB04091;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype B).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
```

PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54485.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 4b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 440 AA;

Query Match 91.9%; Score 2340; DB 4; Length 440;
Best Local Similarity 99.5%; Pred. No. 2.3e-165;
Matches 438; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          33 MFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 92
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MANKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 60

Qy          93 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 152
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy         153 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 212
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 180

Qy         213 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 272
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 240

Qy         273 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 332
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 300

Qy         333 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIK 392
               |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db         301 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDELYNTIQIK 360

Qy         393 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 452
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 420

Qy         453 YNLKLGCNWQFIPKDEGWTE 472
               ||||||||||||||||||||
Db         421 YNLKLGCNWQFIPKDEGWTE 440
```

RESULT 7
AAB04085
ID AAB04085 standard; protein; 439 AA.
XX
AC AAB04085;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin C fragment sequence (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Example 8; Page 38; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 439 AA;

Query Match 91.2%; Score 2322; DB 4; Length 439;
Best Local Similarity 99.3%; Pred. No. 5e-164;
Matches 436; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         34 FNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVT 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVT 60

Qy         94 QNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRII 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIT 120

Qy        154 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIR 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIR 180

Qy        214 EVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPL 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPL 240

Qy        274 MYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSN 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 MYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSN 300

Qy        334 SQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKE 393
              |||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||
Db        301 SQSINDDIVRKEDYIYLDFFNLNQEQRVYTYKKFKKEEEKLFLAPISDSDEFYNTIQIKE 360

Qy        394 YDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPY 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPY 420

Qy        454 NLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||
Db        421 NLKLGCNWQFIPKDEGWTE 439
```

RESULT 8
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 90.6%; Score 2307; DB 2; Length 472;
Best Local Similarity 89.8%; Pred. No. 7.1e-163;
Matches 424; Conservative 26; Mismatches 22; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||:|||||||||||||:|||||||||||||||:|||||||||||||||||:|
Db         61 GYGAKVEVYDGVKLNDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||:|||||||| |||| ||||| ||| |||| ||||||:||||||||||
Db        181 NRWFFVTITNNLDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              |||||:|:||||:| ||||||||||||||||||||||||||||||||||| ||| ||
Db        241 SIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||| |||||||| |||||:||||||||||:|||||||||||||||||:||    ::|||
Db        301 EILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              || |||||::|||||: ||||:||| ||:|||||||:||||||||||||||:||||||||
Db        361 VYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||:: ::|||||||||||||||||||    ||||||||||||||||
Db        421 IHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 472
```

RESULT 9
AAB04082
ID AAB04082 standard; protein; 848 AA.
XX
AC AAB04082;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX

```
PF   12-MAY-2000;.2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Disclosure; Page 13; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 848 AA;

  Query Match             90.3%;  Score 2298;  DB 4;  Length 848;
  Best Local Similarity   98.2%;  Pred. No. 7.1e-162;
  Matches  439;  Conservative    0;  Mismatches    4;  Indels    4;  Gaps    4;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 464

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 NSKIRTTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 524

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 IRG-RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 583

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 643

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSY-IKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 325
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        644 DFWGNPLMYNKEYYMFNAGNKNSYTIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 703

Qy        326 FIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEF 385
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||
Db        704 FIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEE-LFLAPISDSDEF 762

Qy        386 YNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYL 445
              |||||||||||||||||||||||||||||||||||||||||||||| || |||| |||
Db        763 YNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEVKDVFCISDWYL 822

Qy        446 KEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db        823 -EVKRKPYNLKLGCNWQFIPKDEGWTE 848


RESULT 10
AAW56017
ID   AAW56017 standard; protein; 1169 AA.
XX
AC   AAW56017;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH728/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW   detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB002273.
XX
PR   23-AUG-1996;   96GB-00017671.
PR   13-DEC-1996;   96GB-00025996.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Shone CC,  Quinn CP,  Foster KA;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26289.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT   immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS   Example 2; Page 91-94; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from the
CC   present invention. The present invention describes recombinant neurotoxin
CC   proteins which comprise a first and second domain, where the first domain
CC   is adapted to cleave one or more vesicle or plasma-membrane associated
CC   proteins essential to exocytosis, and where the second domain is adapted:
CC   (a) to translocate the protein into a cell; (b) to increase the
CC   solubility of the protein compared to the solubility of the first domain
CC   on its own, or (c) both to translocate the protein into a cell and to
CC   increase the solubility of the protein compared to the solubility of the
CC   first domain on its own, the protein being free of clostridial neurotoxin
CC   (CN) and free of CN precursor that can be converted into toxin by
CC   proteolytic action. The recombinant proteins can be used as therapeutic
CC   agents for targeting cells expressing a relevant substrate. The products
CC   can also be used as immunogens and as non-toxic standards for the
CC   assessment and development of in vitro assays for the detection of
CC   functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC   environmental samples
XX
SQ   Sequence 1169 AA;

  Query Match             66.9%;  Score 1703;  DB 2;  Length 1169;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-117;
  Matches  324;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965
```

```
Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYL 350
              ||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYL 1169
```

RESULT 11
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30596.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC express BotG soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX

SQ Sequence 473 AA;

Query Match 53.8%; Score 1369.5; DB 2; Length 473;
Best Local Similarity 53.3%; Pred. No. 2.9e-93;
Matches 255; Conservative 95; Mismatches 117; Indels 11; Gaps 7;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||::|| | | | :| ||:| |:   ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGAKVEVYDGVELND--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118
              |||| : |   |  ||    |||| :| || |   |:: ::::|:| :||::||:| |||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178
              |: || |: ||||||:|:||:||||:||:||||||||||:| |:||:||||:|:::||:
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237
              |||:|| :||||: | || ||||| |: :  | ::  : :: :| |||    | |:|:|:
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDS 297
              | |:||  ||: : :   | ||| :  |||||||| |: :||:|| | :| |||    :
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        298 PVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLN 356
              :||   |: :|  :  |||::||:| :|||:: |||::| ||:|||: |||||:  |::
Db        301 SMGETAPRTNFNNAA--INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNIS 358

Qy        357 QE-WRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDE 415
              | :|||     |: : :||||||:|   ||: :|||:| |: ||:||:| :||   |
Db        359 DESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKT 415

Qy        416 IGLIGIHRFYES-GIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              || || :| :  | |:: | :|||||:|||: :      |:|||||||| ||||||
Db        416 FGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
```

RESULT 12
AAY77139
ID AAY77139 standard; protein; 449 AA.
XX
AC AAY77139;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype G (BoNTG) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype G; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87217.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum

PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 28; Page 47-48; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 449 AA;

Query Match 47.6%; Score 1210.5; DB 3; Length 449;
Best Local Similarity 50.7%; Pred. No. 1.7e-81;
Matches 230; Conservative 95; Mismatches 118; Indels 11; Gaps 7;

```
Qy         25 MADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKL 82
                | ||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||    ||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         83 TSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSG 142
               :| || |   |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 201
               ||:||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || |||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        202 KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY 261
                |: :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | |||
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        262 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY 321
               :  |||||||||| |: :||:|| | :| |||    : :||   |: :|  :  |||::||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLY 298

Qy        322 IGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPI 379
               :| :|||:: |||::| ||:|||: |||||:  |:: | :|||      |: : :||||||
Db        299 LGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPI 358

Qy        380 SDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYF 438
               :|   ||: :|||:| |: ||:||:| :||   |   || || :| :  | |:: | :||
Db        359 NDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYF 415

Qy        439 CISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
               |||:|||: :      |:|||||||| ||||||
Db        416 CISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 449
```

RESULT 13
AAB04167
ID AAB04167 standard; protein; 449 AA.
XX
AC AAB04167;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype G).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54491.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 10b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 449 AA;

Query Match 47.6%; Score 1210.5; DB 4; Length 449;
Best Local Similarity 50.7%; Pred. No. 1.7e-81;
Matches 230; Conservative 95; Mismatches 118; Indels 11; Gaps 7;

```
Qy         25 MADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKL 82
                | ||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||    ||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         83 TSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSG 142
               :| || |   |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 201
               ||:||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || |||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        202 KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY 261
                |: :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | |||
```

```
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        262 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY 321
              :  ||||||||| |: :||:|| | :| ||| : :||  |: :|  :  |||::||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLY 298

Qy        322 IGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPI 379
              :| :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||
Db        299 LGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPI 358

Qy        380 SDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYF 438
              :|   ||: :|||:| |: ||:||:| :||   |   || || :| :  | |:: | :||
Db        359 NDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYF 415

Qy        439 CISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||:|||: :       |:|||||||| ||||||
Db        416 CISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 449
```

RESULT 14
AAR95009
ID AAR95009 standard; protein; 462 AA.
XX
AC AAR95009;
XX
DT 07-JUL-1996 (first entry)
XX
DE Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .21
FT   /label= Polyhistidine_tag
FT Protein   22. .462
FT   /label= C-fragment
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29246.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 7; Page 340-342; 434pp; English.
XX
CC pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC fragment C (see also AAR95008) of the Clostridium botulinum type A
CC neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC soluble protein and was purified by metal chelate affinity chromatography
CC to obtain a product free of endotoxin contamination that may be useful as
CC an immunogen in vaccine compsns
XX
SQ Sequence 462 AA;

```
  Query Match             38.8%;  Score 988.5;  DB 2;  Length 462;
  Best Local Similarity   42.7%;  Pred. No. 5.5e-65;
  Matches  207;  Conservative   84;  Mismatches  155;  Indels   39;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              |||||||||||||||||||||||||   |:  | :|   |:|  ||||||: |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVELN--DKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118
                | :|: :   |  :  |||| :| :  :||| |     |::||:: :|| |||||||||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178
               |      ::|||||||||:||||||:|:    |||||  |     : | |:|:   :||:
Db        119 FN---SISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237
              ||||| ||||||| |||:||||||:|     | ::  : |:  |:|||||  |  ::||:
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKK-- 295
              |||::|: ||::  |:: |  || |  ||||||: | |:| ||| |  : | |: :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        296 -------DSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYI 348
                    | | ::| : |  :|        || | ||||:: ::     |:||| | :
Db        296 IRGYMYLKGPRGSVMTTNIYLNSS-------LYRGTKFIIKKYASGN--KDNIVRNNDRV 346

Qy        349 YLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP-TYSCQLLFK 407
              |::    |:|:|: |        |: |    | |        : :|  ::|  |  |::
Db        347 YINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--N 404

Qy        408 KDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKD 467
              : : ::|| || |:|                | || ::::|   :  |||:|:||| |
Db        405 LQDNNGNDIGFIGFHQFNNIA--------KLVASNWYNRQIERS--SRTLGCSWEFIPVD 454

Qy        468 EGWTE 472
              :|| |
Db        455 DGWGE 459
```

RESULT 15
AAW68390
ID AAW68390 standard; protein; 462 AA.
XX
AC AAW68390;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism.
XX
OS Clostridium botulinum; serotype A.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .21
FT   /note= "N-terminal histidine tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.

XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30572, AAV30575.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 24; Page 265-267; 428pp; English.
XX
CC This is the amino acid sequence of a histidine-tagged fragment C
CC polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on a poly-histidine binding
CC affinity resin. The invention relates to recombinant proteins derived
CC from C. botulinum toxins. Methods are provided which allow the isolation
CC of soluble recombinant proteins that are free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant proteins are
CC E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are useful
CC in the treatment of humans and animals at risk of intoxication with
CC clostridial toxin
XX
SQ Sequence 462 AA;

Query Match 38.8%; Score 988.5; DB 2; Length 462;
Best Local Similarity 42.7%; Pred. No. 5.5e-65;
Matches 207; Conservative 84; Mismatches 155; Indels 39; Gaps 11;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||    |:  | :|   |:|  ||||||: |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVELN--DKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118
               | :|: :   |  :  |||| :| :  :||| |     |::||:: :|| |||||||||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178
              |      ::|||||||||:||||||:|:    ||||| |     : | |:|:   :||:
Db        119 FN---SISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237
              ||||| ||||||| |||:||||||:|     | ::  : |:  |:|||||  |  ::||:
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKK-- 295
              |||::|: ||::  |:: |  || |  ||||||: | |:| ||| |  : | |: :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        296 -------DSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYI 348
                    | | ::| : |  :|       || | ||||:: ::     |:|||  | :
Db        296 IRGYMYLKGPRGSVMTTNIYLNSS-------LYRGTKFIIKKYASGN--KDNIVRNNDRV 346

Qy        349 YLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP-TYSCQLLFK 407
              |::    |:|:|: |        |: |    | |       : :|  ::|  |  |::
Db        347 YINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--N 404

Qy        408 KDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKD 467
                : : ::|| || |:|                | || ::::|  :  |||:|:||| |
Db        405 LQDNNGNDIGFIGFHQFNNIA--------KLVASNWYNRQIERS--SRTLGCSWEFIPVD 454

Qy        468 EGWTE 472
              :|| |
Db        455 DGWGE 459
```

Search completed: September 10, 2004, 03:33:27
Job time : 33.1347 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 10.0183 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 2545
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1703	66.9	1169	4	US-09-255-829-20	Sequence 20, Appl
2	988.5	38.8	462	1	US-08-480-604A-26	Sequence 26, Appl
3	988.5	38.8	462	2	US-08-405-496A-26	Sequence 26, Appl
4	988.5	38.8	462	3	US-08-915-136-26	Sequence 26, Appl
5	988.5	38.8	462	4	US-09-084-517-26	Sequence 26, Appl
6	838.5	32.9	438	1	US-08-480-604A-23	Sequence 23, Appl
7	838.5	32.9	438	2	US-08-405-496A-23	Sequence 23, Appl
8	838.5	32.9	438	3	US-08-915-136-23	Sequence 23, Appl
9	838.5	32.9	438	4	US-09-084-517-23	Sequence 23, Appl
10	838.5	32.9	1296	1	US-08-480-604A-28	Sequence 28, Appl
11	838.5	32.9	1296	2	US-08-405-496A-28	Sequence 28, Appl
12	838.5	32.9	1296	3	US-08-915-136-28	Sequence 28, Appl
13	838.5	32.9	1296	4	US-09-084-517-28	Sequence 28, Appl
14	591	23.2	452	1	US-07-618-312A-2	Sequence 2, Appli
15	591	23.2	452	1	US-08-280-228-2	Sequence 2, Appli
16	590	23.2	452	1	US-07-618-312A-4	Sequence 4, Appli
17	590	23.2	452	1	US-08-280-228-4	Sequence 4, Appli
18	590	23.2	618	1	US-08-668-381A-5	Sequence 5, Appli
19	590	23.2	853	4	US-08-913-880C-17	Sequence 17, Appl
20	590	23.2	858	4	US-08-913-880C-16	Sequence 16, Appl
21	590	23.2	860	4	US-08-913-880C-15	Sequence 15, Appl
22	590	23.2	862	4	US-08-913-880C-14	Sequence 14, Appl
23	590	23.2	865	4	US-08-913-880C-13	Sequence 13, Appl
24	590	23.2	866	4	US-08-913-880C-12	Sequence 12, Appl
25	590	23.2	874	4	US-08-913-880C-11	Sequence 11, Appl
26	590	23.2	875	4	US-08-913-880C-10	Sequence 10, Appl
27	590	23.2	1315	4	US-08-913-880C-1	Sequence 1, Appli
28	571	22.4	452	1	US-08-110-786A-8	Sequence 8, Appli
29	171.5	6.7	141	4	US-09-465-276-1	Sequence 1, Appli
30	165.5	6.5	140	3	US-08-446-114A-22	Sequence 22, Appl
31	164	6.4	2391	2	US-08-446-855A-2	Sequence 2, Appli
32	164	6.4	2391	3	US-09-150-741-2	Sequence 2, Appli
33	150	5.9	23	1	US-08-480-604A-24	Sequence 24, Appl
34	150	5.9	23	2	US-08-405-496A-24	Sequence 24, Appl
35	150	5.9	23	3	US-08-915-136-24	Sequence 24, Appl
36	150	5.9	23	4	US-09-084-517-24	Sequence 24, Appl
37	147	5.8	323	1	US-08-591-989-5	Sequence 5, Appli
38	147	5.8	803	4	US-09-543-681A-4886	Sequence 4886, Ap
39	146	5.7	22	4	US-09-660-742-3	Sequence 3, Appli
40	146	5.7	53	2	US-08-651-818A-19	Sequence 19, Appl
41	146	5.7	53	3	US-09-184-826-19	Sequence 19, Appl
42	146	5.7	54	2	US-08-651-818A-23	Sequence 23, Appl
43	146	5.7	54	3	US-09-184-826-23	Sequence 23, Appl
44	143.5	5.6	476	3	US-09-316-083-3	Sequence 3, Appli
45	143.5	5.6	476	4	US-09-933-700-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 66.9%; Score 1703; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 8.6e-122;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYL 350
              ||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYL 1169


RESULT 2
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match            38.8%;  Score 988.5;  DB 1;  Length 462;
  Best Local Similarity  42.7%;  Pred. No. 1e-67;
  Matches  207;  Conservative   84;  Mismatches  155;  Indels   39;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||    |: | :|   |:| ||||||: |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVELN--DKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118
               | :|: :   |  :  |||| :| :  :||| |     |::||:: :|| |||||||||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178
               |      ::|||||||||:||||||:|:    |||||  |     : | |:|:   :||:
Db        119 FN---SISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237
              |||||  |||||||  |||:||||||:|    | ::  : |:  |:|||||  |  ::||:
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKK-- 295
              |||::|: ||::  |:: |  || |  ||||||: | |:| ||| |  : | |: :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        296 -------DSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYI 348
                    | | ::| : |  :|       || | ||||:: ::     |:||| | :
Db        296 IRGYMYLKGPRGSVMTTNIYLNSS-------LYRGTKFIIKKYASGN--KDNIVRNNDRV 346

Qy        349 YLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP-TYSCQLLFK 407
              |::    |:|:|: |        |: |    | |       : :|  ::|  |  |::
Db        347 YINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--N 404

Qy        408 KDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKD 467
              : : ::|| || |:|                | || ::::|   :  |||:|:||| |
Db        405 LQDNNGNDIGFIGFHQFNNIA--------KLVASNWYNRQIERS--SRTLGCSWEFIPVD 454

Qy        468 EGWTE 472
              :|| |
Db        455 DGWGE 459


RESULT 3
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/405,496A
;       FILING DATE:  16-MAR-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  INGOLIA, DIANE E.
;       REGISTRATION NUMBER:  40,027
;       REFERENCE/DOCKET NUMBER:  OPHD-01308
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  462 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match             38.8%;  Score 988.5;  DB 2;  Length 462;
  Best Local Similarity   42.7%;  Pred. No. 1e-67;
  Matches  207;  Conservative   84;  Mismatches  155;  Indels   39;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||    |:  | :|   |:|  ||||||: |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVELN--DKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118
              | :|: :   |  :  |||| :| :  :||| |     |::||:: :|| |||||||||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178
              |      ::|||||||||:||||||:|:    |||||  |     :  | |:|:   :||:
Db        119 FN---SISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237
              ||||| ||||||| |||:||||||:|     | ::  : |:  |:|||||  |  ::||:
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKK-- 295
              |||::|: ||::  |:: |  || |  ||||||: | |:| ||| |  : | |: :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        296 -------DSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYI 348
                    | | ::| : |  :|       || | ||||:: ::     |:|||  | :
Db        296 IRGYMYLKGPRGSVMTTNIYLNSS-------LYRGTKFIIKKYASGN--KDNIVRNNDRV 346

Qy        349 YLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP-TYSCQLLFK 407
              |::    |:|:|: |       |: |    | |       : :|  ::|  |  |::
Db        347 YINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--N 404

Qy        408 KDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKD 467
              : : ::|| || |:|               | || ::::|   :  |||:|:||| |
Db        405 LQDNNGNDIGFIGFHQFNNIA--------KLVASNWYNRQIERS--SRTLGCSWEFIPVD 454

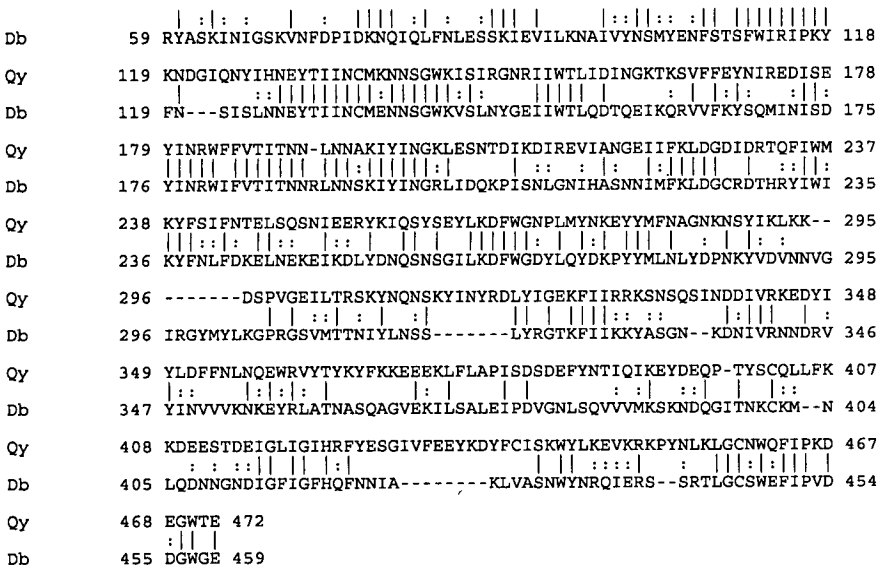
Qy        468 EGWTE 472
              :|| |
Db        455 DGWGE 459
```

```
RESULT 4
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             38.8%;  Score 988.5;  DB 3;  Length 462;
  Best Local Similarity   42.7%;  Pred. No. 1e-67;
  Matches  207;  Conservative   84;  Mismatches  155;  Indels   39;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||    |:  | :|   |:|  ||||||: |:|||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVELN--DKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118
```



```
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178

Db        119 FN---SISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237

Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKK-- 295

Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        296 -------DSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYI 348

Db        296 IRGYMYLKGPRGSVMTTNIYLNSS-------LYRGTKFIIKKYASGN--KDNIVRNNDRV 346

Qy        349 YLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP-TYSCQLLFK 407

Db        347 YINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--N 404

Qy        408 KDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKD 467

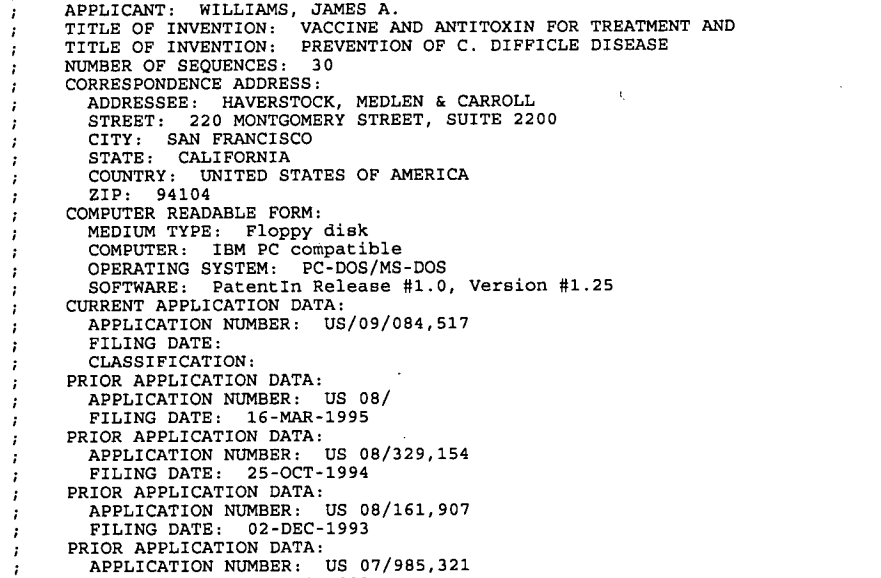
Db        405 LQDNNGNDIGFIGFHQFNNIA--------KLVASNWYNRQIERS--SRTLGCSWEFIPVD 454

Qy        468 EGWTE 472

Db        455 DGWGE 459
```

RESULT 5
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-26

Query Match 38.8%; Score 988.5; DB 4; Length 462;
Best Local Similarity 42.7%; Pred. No. 1e-67;
Matches 207; Conservative 84; Mismatches 155; Indels 39; Gaps 11;



```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGAKVEVYDGVELN--DKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118

Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178

Db        119 FN---SISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237

Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKK-- 295

Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        296 -------DSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYI 348

Db        296 IRGYMYLKGPRGSVMTTNIYLNSS-------LYRGTKFIIKKYASGN--KDNIVRNNDRV 346

Qy        349 YLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP-TYSCQLLFK 407

Db        347 YINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKM--N 404

Qy        408 KDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKD 467

Db        405 LQDNNGNDIGFIGFHQFNNIA--------KLVASNWYNRQIERS--SRTLGCSWEFIPVD 454

Qy        468 EGWTE 472

Db        455 DGWGE 459
```

RESULT 6
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO

```
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,604A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/422,711
;       FILING DATE:  14-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/405,496
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  INGOLIA, DIANE E.
;       REGISTRATION NUMBER:  40,027
;       REFERENCE/DOCKET NUMBER:  OPHD-01763
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  23:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  438 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match             32.9%;  Score 838.5;  DB 1;  Length 438;
  Best Local Similarity   39.7%;  Pred. No. 2.8e-56;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::|||||||:|||||:|:
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 120

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
              ||||| |     : | |:|:   :||:||||| |||||| |||:||||||:|
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  : |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :              | | ::| : |  :|
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 295

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|:::    |:|:|: |        |: |
Db        296 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 351

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db        352 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 401

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::::|   :  |||:|:||| |:|| |
Db        402 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 7
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match             32.9%;  Score 838.5;  DB 2;  Length 438;
  Best Local Similarity   39.7%;  Pred. No. 2.8e-56;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::|||||||:|||||:|:
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 120
```

```
Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 295

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
Db        296 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 351

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
Db        352 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 401

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
Db        402 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 8
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match            32.9%;  Score 838.5;  DB 3;  Length 438;
  Best Local Similarity  39.7%;  Pred. No. 2.8e-56;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 120

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 295

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
Db        296 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 351

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
Db        352 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 401

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
Db        402 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 9
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-23

  Query Match             32.9%;  Score 838.5;  DB 4;  Length 438;
  Best Local Similarity   39.7%;  Pred. No. 2.8e-56;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|    |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::||||||||:||||||:|:
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 120

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
              ||||| |     : | |:|:   :||:||||| ||||||| |||:||||||:|
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  : |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :              | | ::| : |  :|
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 295

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:||| | :|::    |:|:|: |       |: |
Db        296 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 351

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db        352 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 401

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::::|   :  |||:|:||| |:|| |
Db        402 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 435


RESULT 10
US-08-480-604A-28
```

```
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             32.9%;  Score 838.5;  DB 1;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 1.2e-55;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|    |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::||||||||:||||||:|:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
```

```
                    |||||  |     : |  |:|:   :||:|||||  |||||||  |||:||||||:|
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  :  |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :              | | ::| : |  :|
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 1153

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |        |: |
Db       1154 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1209

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db       1210 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 1259

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              | || ::::|   :  |||:|:||| |:|| |
Db       1260 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 11
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match           32.9%;  Score 838.5;  DB 2;  Length 1296;
  Best Local Similarity  39.7%;  Pred. No. 1.2e-55;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:|||| | :|: :   |  :  |||| :| :  :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::|||||||||:||||||:|:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
              |||||  |     : | |:|:   :||:|||||  |||||||  |||:||||||:|
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  :  |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :              | | ::| : |  :|
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 1153

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |        |: |
Db       1154 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1209

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db       1210 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 1259

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              | || ::::|   :  |||:|:||| |:|| |
Db       1260 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 12
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIÀ, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             32.9%;  Score 838.5;  DB 3;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 1.2e-55;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::|||||||||:||||||:|:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
                |||| |     : | |:|:   :||:|||| ||||||| |||:||||||:|
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  : |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :             | | ::| : |  :|
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 1153

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |        |: |
Db       1154 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1209

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db       1210 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 1259

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::::|   :  |||:|:||| |:|| |
Db       1260 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 13
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             32.9%;  Score 838.5;  DB 4;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 1.2e-55;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::|||||||||:||||||:|:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
                |||| |     : | |:|:   :||:|||| ||||||| |||:||||||:|
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  : |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :             | | ::| : |  :|
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 1153
```

```
Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |       |: |
Db       1154 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1209

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db       1210 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 1259

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                 | || ::::|   :  |||:|:||| |:|| |
Db       1260 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 14
US-07-618-312A-2
; Sequence 2, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             23.2%;  Score 591;  DB 1;  Length 452;
  Best Local Similarity   32.0%;  Pred. No. 2.4e-37;
  Matches  141;  Conservative   88;  Mismatches  184;  Indels   28;  Gaps   12;

Qy         46 ILNLRYKDNNLIDLSGYGAKVEVYDGVEL----NDKNQFKLTSSANSKIRVTQNQNIIFN 101
              ||||   ::  ||:||: : |  |   :|    | |    | :: :|:: | :  :| :|
Db         23 ILNLDINNDINIDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 81

Qy        102 SVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
               :| :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :||||
Db         82 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 141

Qy        157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LNNAKIYINGKLESNTDIKDIRE 214
                |  |: : : |  :: :  : |: |:| |:|||: |::| :|||| |  : :|  :
Db        142 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 200

Qy        215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
              :  :  | |||   :  |:: :  | ||   |:   ||: |         :|:|||||||
Db        201 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 260

Qy        275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
              |: |||:    : :  ::||  :    :     |       | || || | ||||:| : :
Db        261 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 320

Qy        335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTY----KYFKKEEEKLFLAPISDSDEFYNTIQ 390
                | |  |:  |:| | : : |    :  |       |   :  | :   :     | ::
Db        321 NEI-DSFVKSGDFIKL-YVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 378

Qy        391 IKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKR 450
              :  :  ||| ||    |: ::  :||:| |     | :  :|  | ||   :|
Db        379 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 433

Qy        451 KPYNLKLGCNWQFIPKDEGWT 471
              |     |||:| |:| |||||
Db        434 K----ILGCDWYFVPTDEGWT 450


RESULT 15
US-08-280-228-2
; Sequence 2, Application US/08280228
; Patent No. 5571694
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. 5571694th Glebe Road
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,228
;      FILING DATE:  25-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/618,312
;      FILING DATE:  27-NOV-1990
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mary J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  117-163
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
```

; LENGTH: 452 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-280-228-2

Query Match 23.2%; Score 591; DB 1; Length 452;
Best Local Similarity 32.0%; Pred. No. 2.4e-37;
Matches 141; Conservative 88; Mismatches 184; Indels 28; Gaps 12;

```
Qy          46 ILNLRYKDNNLIDLSGYGAKVEVYDGVEL----NDKNQFKLTSSANSKIRVTQNQNIIFN 101
               ||||   ::  ||:||: : |  |   :|    | |    | :: :|:: | :  :| :|
Db          23 ILNLDINNDINIDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 81

Qy         102 SVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
                :| :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :||||
Db          82 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 141

Qy         157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LNNAKIYINGKLESNTDIKDIRE 214
               |  |: : : |  :: :  : |: |:| |:|||: |::| :|||| |  : :|  :
Db         142 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 200

Qy         215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
               :  :  |  |||   :  |:: :  | ||   |:   ||: |       :|:|||||
Db         201 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 260

Qy         275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
               |: |||:    : :  ::|| :    :      |       | || || | ||||:| : :
Db         261 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 320

Qy         335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTY----KYFKKEEEKLFLAPISDSDEFYNTIQ 390
                 | |  |:  |:| | : : |    :  |       |   :  | :   :     |  ::
Db         321 NEI-DSFVKSGDFIKL-YVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 378

Qy         391 IKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKR 450
                 : :  ||| ||    |: ::  :||:| |       |  :  :|    | ||   :|
Db         379 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 433

Qy         451 KPYNLKLGCNWQFIPKDEGWT 471
               |     |||:| |:| |||||
Db         434 K----ILGCDWYFVPTDEGWT 450
```

Search completed: September 10, 2004, 03:46:56
Job time : 12.0183 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 91.3645 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 2545
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2545	100.0	472	12	US-10-272-898-46	Sequence 46, Appl
2	2545	100.0	472	15	US-10-354-774-46	Sequence 46, Appl
3	2545	100.0	472	15	US-10-271-012-46	Sequence 46, Appl
4	2545	100.0	472	16	US-10-729-122-46	Sequence 46, Appl
5	2545	100.0	472	16	US-10-729-039-46	Sequence 46, Appl
6	2382	93.6	1290	15	US-10-360-101-220	Sequence 220, App
7	2382	93.6	1290	15	US-10-452-024-119	Sequence 119, App
8	2382	93.6	1291	12	US-10-272-898-42	Sequence 42, Appl
9	2382	93.6	1291	15	US-10-354-774-42	Sequence 42, Appl
10	2382	93.6	1291	15	US-10-271-012-42	Sequence 42, Appl
11	2382	93.6	1291	15	US-10-452-024-2	Sequence 2, Appli
12	2382	93.6	1291	15	US-10-452-024-118	Sequence 118, App
13	2382	93.6	1291	16	US-10-729-122-42	Sequence 42, Appl
14	2382	93.6	1291	16	US-10-729-039-42	Sequence 42, Appl
15	2340	91.9	440	10	US-09-910-186A-8	Sequence 8, Appli
16	2307	90.6	472	12	US-10-272-898-44	Sequence 44, Appl
17	2307	90.6	472	15	US-10-354-774-44	Sequence 44, Appl
18	2307	90.6	472	15	US-10-271-012-44	Sequence 44, Appl
19	2307	90.6	472	16	US-10-729-122-44	Sequence 44, Appl
20	2307	90.6	472	16	US-10-729-039-44	Sequence 44, Appl
21	2267	89.1	1291	15	US-10-452-024-122	Sequence 122, App
22	2265	89.0	1291	15	US-10-452-024-121	Sequence 121, App
23	2259	88.8	1291	15	US-10-452-024-123	Sequence 123, App
24	2188	86.0	1291	15	US-10-452-024-142	Sequence 142, App
25	2133	83.8	441	15	US-10-452-024-188	Sequence 188, App
26	2125	83.5	1291	12	US-10-272-898-40	Sequence 40, Appl
27	2125	83.5	1291	15	US-10-354-774-40	Sequence 40, Appl
28	2125	83.5	1291	15	US-10-271-012-40	Sequence 40, Appl
29	2125	83.5	1291	15	US-10-452-024-117	Sequence 117, App
30	2125	83.5	1291	15	US-10-205-516-4	Sequence 4, Appli
31	2125	83.5	1291	16	US-10-729-122-40	Sequence 40, Appl
32	2125	83.5	1291	16	US-10-729-039-40	Sequence 40, Appl
33	2125	83.5	1301	15	US-10-205-516-18	Sequence 18, Appl
34	1703	66.9	1169	14	US-10-241-596-20	Sequence 20, Appl
35	1369.5	53.8	473	12	US-10-272-898-79	Sequence 79, Appl
36	1369.5	53.8	473	15	US-10-354-774-79	Sequence 79, Appl
37	1369.5	53.8	473	15	US-10-271-012-79	Sequence 79, Appl
38	1369.5	53.8	473	16	US-10-729-122-79	Sequence 79, Appl
39	1369.5	53.8	473	16	US-10-729-039-79	Sequence 79, Appl
40	1210.5	47.6	449	10	US-09-910-186A-18	Sequence 18, Appl
41	1206.5	47.4	1297	12	US-10-272-898-77	Sequence 77, Appl
42	1206.5	47.4	1297	15	US-10-354-774-77	Sequence 77, Appl
43	1206.5	47.4	1297	15	US-10-271-012-77	Sequence 77, Appl
44	1206.5	47.4	1297	15	US-10-452-024-7	Sequence 7, Appli
45	1206.5	47.4	1297	15	US-10-452-024-150	Sequence 150, App

ALIGNMENTS

```
RESULT 1
US-10-272-898-46
; Sequence 46, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
```

```
;           LENGTH: 472 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-272-898-46

  Query Match             100.0%;  Score 2545;  DB 12;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-189;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 2
US-10-354-774-46
; Sequence 46, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-354-774-46

  Query Match             100.0%;  Score 2545;  DB 15;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-189;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 3
US-10-271-012-46
; Sequence 46, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-271-012-46

  Query Match           100.0%;  Score 2545;  DB 15;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-189;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 4
US-10-729-122-46
; Sequence 46, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 472 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-729-122-46

  Query Match           100.0%;  Score 2545;  DB 16;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-189;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 5
US-10-729-039-46
```

```
; Sequence 46, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 472 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-729-039-46

  Query Match            100.0%;  Score 2545;  DB 16;  Length 472;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-189;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKN 120

Qy        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYI 180

Qy        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYF 240

Qy        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG 300

Qy        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWR 360

Qy        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 420

Qy        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472


RESULT 6
US-10-360-101-220
; Sequence 220, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
;  APPLICANT: Moll, Gert  N.
;  APPLICANT:  Leenhouts, Cornelis J.
;  TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
;  FILE REFERENCE: 2183-5673
;  CURRENT APPLICATION NUMBER: US/10/360,101
;  CURRENT FILING DATE:  2003-02-07
;  PRIOR APPLICATION NUMBER: EP 02077060.8
;  PRIOR FILING DATE: 2002-05-24
;  NUMBER OF SEQ ID NOS: 309
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 220
;   LENGTH: 1290
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: sequence of clostridium botulinum toxin type B
US-10-360-101-220

  Query Match            93.6%;  Score 2382;  DB 15;  Length 1290;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 7
US-10-452-024-119
; Sequence 119, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
```

```
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 119
;   LENGTH: 1290
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-119

  Query Match             93.6%;  Score 2382;  DB 15;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 8
US-10-272-898-42
; Sequence 42, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-272-898-42

  Query Match             93.6%;  Score 2382;  DB 12;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 9
US-10-354-774-42
; Sequence 42, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
```

```
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1291 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-354-774-42

  Query Match           93.6%;  Score 2382;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 10
US-10-271-012-42
;
; Sequence 42, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/271,012
;            FILING DATE: 15-Oct-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 42:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1291 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-271-012-42

  Query Match           93.6%;  Score 2382;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 11
US-10-452-024-2
; Sequence 2, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-2

  Query Match            93.6%;  Score 2382;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 12
US-10-452-024-118
; Sequence 118, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 118
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-118

  Query Match            93.6%;  Score 2382;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||||||||
Db       1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 13
US-10-729-122-42
; Sequence 42, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;          CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;          PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;          ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;          TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;       INFORMATION FOR SEQ ID NO: 42:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-729-122-42

  Query Match             93.6%;  Score 2382;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy          87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy         147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy         207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy         267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy         327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy         387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy         447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
               |||||||||||||||||||||||||
Db        1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 14
US-10-729-039-42
; Sequence 42, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;      INFORMATION FOR SEQ ID NO: 42:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-729-039-42

  Query Match             93.6%;  Score 2382;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-176;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy          87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy         147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy         207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy         267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy         327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy         387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy         447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
               |||||||||||||||||||||||||
Db        1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 15
US-09-910-186A-8
; Sequence 8, Application US/09910186A
```

```
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 440
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
;   OTHER INFORMATION: sequence
US-09-910-186A-8

  Query Match          91.9%;  Score 2340;  DB 10;  Length 440;
  Best Local Similarity  99.5%;  Pred. No. 1.9e-173;
  Matches  438;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         33 MFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 92
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MANKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRV 60

Qy         93 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy        153 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDI 180

Qy        213 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 REVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNP 240

Qy        273 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKS 300

Qy        333 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIK 392
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        301 NSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDELYNTIQIK 360

Qy        393 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKRKP 420

Qy        453 YNLKLGCNWQFIPKDEGWTE 472
              ||||||||||||||||||||
Db        421 YNLKLGCNWQFIPKDEGWTE 440
```

Search completed: September 10, 2004, 04:58:00
Job time : 92.3645 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 8.33859 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 2545
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2382	93.6	1291	1	A48940	bontoxilysin (EC 3
2	2125	83.5	1291	2	I40631	non-proteolytic bo
3	1206.5	47.4	1297	2	S39791	neurotoxin - Clost
4	838.5	32.9	1296	1	BTCLAB	bontoxilysin (EC 3
5	793	31.2	1268	2	S33411	botulinum neurotox
6	784.5	30.8	1296	2	I40645	botulinum neurotox
7	740	29.1	1274	2	I40813	neurotoxin type F
8	713.5	28.0	1252	2	S21178	botulinum neurotox
9	697	27.4	1251	2	JH0256	botulinum neurotox
10	621.5	24.4	1276	2	S11455	botulinum neurotox
11	590	23.2	1315	1	BTCLTN	tentoxilysin (EC 3
12	564	22.2	1285	2	S70582	botulinum neurotox
13	550	21.6	1291	2	S46431	botulinum neurotox
14	550	21.6	1291	2	A49777	botulinum neurotox
15	349	13.7	366	2	S48110	neurotoxin type F
16	328	12.9	369	2	S48109	neurotoxin type F
17	327.5	12.9	367	2	S48106	neurotoxin type E
18	238.5	9.4	1162	2	A47708	progenitor toxin n
19	234.5	9.2	1162	2	I40817	botulinum toxin no
20	226	8.9	1196	2	JQ1467	toxin, nontoxic co
21	226	8.9	1196	2	S46430	botulinum neurotox
22	223	8.8	1165	2	I40644	botulinum neurotox
23	213.5	8.4	1193	2	S68218	botulinum neurotox
24	197.5	7.8	1193	2	JC4901	nontoxic-nonhemagg
25	178	7.0	4981	2	T18489	hypothetical prote
26	176.5	6.9	2485	1	H71621	serine/threonine-s
27	174	6.8	1516	2	E71619	RAD2 endonuclease
28	167	6.6	2013	2	C71610	probable membrane
29	166	6.5	1553	2	T18502	hypothetical prote
30	164	6.4	2391	2	T18410	carbamoyl-phosphat
31	163	6.4	3119	2	T18414	protein g377 - mal
32	163	6.4	3973	2	B71612	hypothetical prote
33	161	6.3	765	2	S72278	ATP-dependent Clp
34	160.5	6.3	763	2	S17998	gene COX1 intron 4
35	160.5	6.3	960	2	S72284	DNA-directed RNA p
36	159.5	6.3	963	2	C90535	conserved hypothet
37	159	6.2	789	2	G90587	lipoprotein [impor
38	157.5	6.2	1061	2	D98008	conserved hypothet
39	157.5	6.2	1247	2	E71616	hypothetical prote
40	157.5	6.2	2136	2	A05037	hypothetical prote
41	157	6.2	971	2	C82880	conserved hypothet
42	156	6.1	1166	2	H71609	hypothetical prote
43	155.5	6.1	2380	2	E71604	hypothetical prote
44	154.5	6.1	437	2	T28180	hypothetical prote
45	154.5	6.1	714	2	C90100	hypothetical prote

ALIGNMENTS

RESULT 1
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisio
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein

A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 93.6%; Score 2382; DB 1; Length 1291;
Best Local Similarity 100.0%; Pred. No. 9.1e-124;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 905

Qy      87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy     147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1025

Qy     207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1085

Qy     267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy     327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1205

Qy     387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1265

Qy     447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
           ||||||||||||||||||||||||||
Db    1266 EVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 2
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic si
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 83.5%; Score 2125; DB 2; Length 1291;
Best Local Similarity 88.7%; Pred. No. 1.3e-109;
Matches 394; Conservative 28; Mismatches 22; Indels 0; Gaps 0;

```
Qy      29 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANS 88
           |||::|||||||||||||||:|||||||||||||||||||||||||||||||||||:|
Db     848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy      89 KIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIR 148
           ||||||||||||:|||||||||||||:|| ||||||||||||||||||||||||||||||
Db     908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy     149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTD 208
           ||||||||||||||||||||||||||||||||||||||||||||:|||||||| |||| |
Db     968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy     209 IKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDF 268
           |||| ||| |||| |||||||:|||||||||||||||:|:|||||:| ||||||||||||
Db    1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy     269 WGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFII 328
           |||||||||||||||||||||||| ||| ||||| |||||||| ||||||||||||||||
Db    1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy     329 RRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNT 388
           ||:|||||||||||||:|| ::||||| ||||||:|||||: ||||:||| |
Db    1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy     389 IQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEV 448
           |:||||||||:|||||||||||||||:|||||||||||:: ::|||||||||||||||
```

```
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||   |||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 3
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             47.4%;  Score 1206.5;  DB 2;  Length 1297;
  Best Local Similarity   50.7%;  Pred. No. 5.8e-59;
  Matches  229;  Conservative   95;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
              |||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||    |||| :
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 144
              | || |   |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 203
              :||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        204 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 263
              : :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 323
              |||||||||  |: :|||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 1148

Qy        324 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 381
              :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||:|
Db       1149 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 1208

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 440
              ||: :|||:| |: ||:||:| :||   |   || || :| :  | |:: | :||||
Db       1209 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 1265

Qy        441 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |:|||: :      |:|||||||| ||||||
Db       1266 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 4
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin,
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ext
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arour
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; N
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prote
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

  Query Match             32.9%;  Score 838.5;  DB 1;  Length 1296;
```

```
  Best Local Similarity   39.7%;  Pred. No. 1.1e-38;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:||||| | :|: :    |  :  |||| :| :  :
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |      ::|||||||:||||||:|:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
              |||||  |      : | |:|:   :||:|||||  |||||||  |||:||||||:|
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  : |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :              | | ::| : |  :|
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 1153

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |        |: |
Db       1154 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1209

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |      : :|  ::|  |  |::      : : ::|| || |:|
Db       1210 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 1259

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::::|   :  |||:|:||| |:|| |
Db       1260 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1293


RESULT 5
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            31.2%;  Score 793;  DB 2;  Length 1268;
  Best Local Similarity   39.2%;  Pred. No. 3.5e-36;
  Matches  177;  Conservative   93;  Mismatches  150;  Indels   32;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | ||| :  ::   |::: |||::|::|  || ||||: : :   : :   ::||| : |
Db        843 DKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYS 902

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: :|||  ||:|| : :||||||:||||| |  ::| ::|||||||||: |||||
Db        903 SRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN--LKN-LNNEYTIINCMRNNNSGW 959

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
              |||:  | ||||| |  |  : : | |    |||:|||:| ||||||| | ::|:||||
Db        960 KISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGN 1019

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
              |     | ::  :  :  |:||: |  | |::: ::|| ||| || :: ||  |  :  |
Db       1020 LTDQKSILNLGNIHVDDNILFKIVGCND-TRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEI-LTRSKYNQNSKYINYRDLY 321
              ||||||| |:|||:||: |     |  : : |:| :  |    |  |:: : : | | ||
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPN--MSVTKNSDILNINRQRGIYSKTNIFSNAR-LY 1135

Qy        322 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPIS 380
              | : |||:  :: : | |: ||| | :|::  : | |:::|        |: : |  ||
Db       1136 TGVEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRIS 1195

Qy        381 DSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCI 440
              :|:  ||: |:   |     :| : || :  : ::|||:| |             :
Db       1196 NSN--YNSNQMIIMDSIGD-NCTMNFKTN--NGNDIGLLGFH-----------LNNLVA 1238

Qy        441 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              | || | ::   |   || | || |: || |
Db       1239 SSWYYKNIRNNTRN--NGCFWSFISKEHGWQE 1268


RESULT 6
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A a
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            30.8%;  Score 784.5;  DB 2;  Length 1296;
  Best Local Similarity   37.9%;  Pred. No. 1.1e-35;
  Matches  173;  Conservative   89;  Mismatches  157;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||:: || ::|||||  ||||: : | |  :  |||| || :  :
Db        862 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 921

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              | | |     |::||:: :|| ||||:|||| :   :  ::|||||||::|||||:|:
Db        922 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSL 978

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
              |||||  |      : | |:|:   :||:|||||  |||||||  |  :|||||:|
Db        979 NYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQ 1038

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              | ::  : |: :|:|||||  |  ::| :|||::|: ||::  |:: |  || |  ||
Db       1039 KPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1098

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              |||||  | |:| ||| |  : | |: :              | | ::| : |  ::
Db       1099 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNST----- 1153

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |        |: |
Db       1154 --LYEGTKFIIKKYASGN--EDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1209

Qy        378 PISDSDEFYNTIQIKEYDEQPTYS-CQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |      : :|  |:|   : |::     : : ::|| || | :        :
Db       1210 EIPDVGNLSQVVVMKSKDDQGIRNKCKM--NLQDNNGNDIGFIGFHLY--------DNIA 1259

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::| :   :   ||:|:||| |:|| |
Db       1260 KLVASNWYNRQVGKA--SRTFGCSWEFIPVDDGWGE 1293
```

```
RESULT 7
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            29.1%;  Score 740;  DB 2;  Length 1274;
  Best Local Similarity  37.0%;  Pred. No. 2.9e-33;
  Matches  171;  Conservative   82;  Mismatches  155;  Indels   54;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              |  |||  ||:   :| :: ||::||::|  ||:||||: : :   | :   ::||| : :
Db        851 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 910

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGW 143
              |  |:: : || :||:|| : :||:|||:||||:      |    |||||||  |||||
Db        911 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGW 967

Qy        144 KISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYI 199
              |||:|  |   ||||| | :|  ::: | |     || |||:| ||||||| | |::|||
Db        968 KISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYI 1027

Qy        200 NGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQ 259
              || |     | :: ::  :  |:||: |  | | :: ::|| :||||| :: ||  |  :
Db       1028 NGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNE 1086

Qy        260 SYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRD 319
                ||::||| |:|||:||:||   |: || |  :| :  |  :    : | ::||:
Db       1087 PDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK- 1143

Qy        320 LYIGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLF--- 375
              || | : |||:       | |: ||| |  |::  :   |:|:|     | |:||:
Db       1144 LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIRTS 1201

Qy        376 -----LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV 430
                  |  |   |  |               :| : |: :  |   |||:| |
Db       1202 NLNDSLGQIIVMDSIGN------------NCTMNFQNNNGS--NIGLLGFH-------- 1238

Qy        431 FEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                :   | ||   ::|      || |  | |: || |
Db       1239 ----SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 1274


RESULT 8
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match            28.0%;  Score 713.5;  DB 2;  Length 1252;
  Best Local Similarity  35.7%;  Pred. No. 8.3e-32;
  Matches  161;  Conservative   91;  Mismatches  162;  Indels   37;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              |  |||  |||:   | :: :||:|||::  :| ||| : : :   |     :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
                |:: ::||  ||::: : :||:||||:||| | |  :   ::|||||||||: |||||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
              |:|:  | ||||| |  |  : : | |     ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
              |     | ::  :  :  |:||:  :   |::| ::||:||: || :: |:  |  :  :
Db       1011 LIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
              ||||||| |:|:|||||: |    |::|  :|||      | |  |  |  :    ||
```

```
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS------TLSINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTYKYFKKEEEKLFLAPISD 381
              | |  |:| :|| | ||::||| | :|::|  :    : :|       :|: :    ||
Db       1124 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTI---KISS 1179

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS 441
              |   :| : :          :| : || :  : : |||:|           |     |
Db       1180 SGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1222

Qy        442 KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
               ||   :  : :    || | || :: || |
Db       1223 TWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251


RESULT 9
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match             27.4%;  Score 697;  DB 2;  Length 1251;
  Best Local Similarity   35.3%;  Pred. No. 6.7e-31;
  Matches  159;  Conservative   91;  Mismatches  163;  Indels   38;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:|||::  :| ||| : : :   |     :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
                 |:: ::||   ||::: : :||:|||:||| | |   :   ::|||||||||: |||||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
              |:|:  | ||||| | :|   : : | |     ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
              |      | ::  :  :  |:||:  :   |::| ::||:||: || :: |:  |  :  :
Db       1011 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
              ||||||| |:|:|||||: |    |::|  : ||      | |  |  |  :    ||
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS------TLSINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTYKYFKKEEEKLFLAPISD 381
              | |  |:| :|| | ||::||| | :|::|  :      :|       :|: :    ||
Db       1124 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTI---KISS 1179

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS 441
              |   :| : :          :| : || :  : : |||:|           |     |
Db       1180 SGNRFNQVVVM----NSVGNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1221

Qy        442 KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
               ||   ::    :   |  | || :: || |
Db       1222 TWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1250


RESULT 10
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Krieglst
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type I
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             24.4%;  Score 621.5;  DB 2;  Length 1276;
  Best Local Similarity   33.9%;  Pred. No. 1e-26;
  Matches  160;  Conservative   78;  Mismatches  163;  Indels   71;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
              :::| :: |:| : | :: ||:|: | | |:| ||| |:| | | |:||    | |||:|
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 144
              | : || |  | ||:::::: : ||||||:| |   :      |||||||| :: |||||
Db        910 SGD-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWK 963

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 203
              : ||   | | | |:| | ||: |:|:     : | |:|||||||| :   |:||||:|
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        204 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 263
              : :  |:|: ||  :  |:| :| :||  | :|:: |:||: |||  :|   |: |
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 323
              :||:||||| :: |||: |    : ||     :|  :|   :|   ||      || |
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVQYPDRSK------LYTG 1132

Qy        324 EKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSD 383
              |:  |:    :   :   | | |     :::: :                  | |:|
Db       1133 NPITIKSVSDKNPYSR--ILNGDNIILHMLYNSRKYMI----------------IRDTD 1173

Qy        384 EFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV----------F 431
              | | |  |  :   |: :|           :|  ||  |    ||           |
Db       1174 TIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSF 1224

Qy        432 EE--------YKDYFCISKWYLKEVKRKPYNLKL---GCNWQFIPKDEGWTE 472
              |         || :    |    |  ||      |:|| :| || |
Db       1225 RENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 11
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
```

```
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglic
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match            23.2%;  Score 590;  DB 1;  Length 1315;
  Best Local Similarity  31.7%;  Pred. No. 5.7e-25;
  Matches  140;  Conservative   89;  Mismatches  184;  Indels   28;  Gaps   12;

Qy         46 ILNLRYKDNNLIDLSGYGAKVEVYDGVEL----NDKNQFKLTSSANSKIRVTQNQNIIFN 101
              ||||   :: : |:||: : |  |   :|    | |    | :: :|:: | :  :| :|
Db        886 ILNLDINNDIISDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 944

Qy        102 SVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
              :| :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :||||
Db        945 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 1004

Qy        157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LNNAKIYINGKLESNTDIKDIRE 214
              |  |: : : |  :: :  : |: |:| |:||||: |::| :|||| |  : :|  :
Db       1005 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 1063

Qy        215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
              :  :  | |||   :  |:: :  | ||   |:   ||: |     :|:|||||||
Db       1064 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 1123

Qy        275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
              |: |||:    : :  ::||  :    :       |       | || || | ||||:| : :
Db       1124 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 1183

Qy        335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTY----KYFKKEEEKLFLAPISDSDEFYNTIQ 390
               | |  |:  |:| | : : |    :  |       |   :  | :   :      | ::
Db       1184 NEI-DSFVKSGDFIKL-YVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 1241

Qy        391 IKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKR 450
               :  :  ||| ||    |: ::  :||:| |       |  :  :|    | ||   :|
Db       1242 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 1296

Qy        451 KPYNLKLGCNWQFIPKDEGWT 471
               |     |||:| |:| |||||
Db       1297 K----ILGCDWYFVPTDEGWT 1313

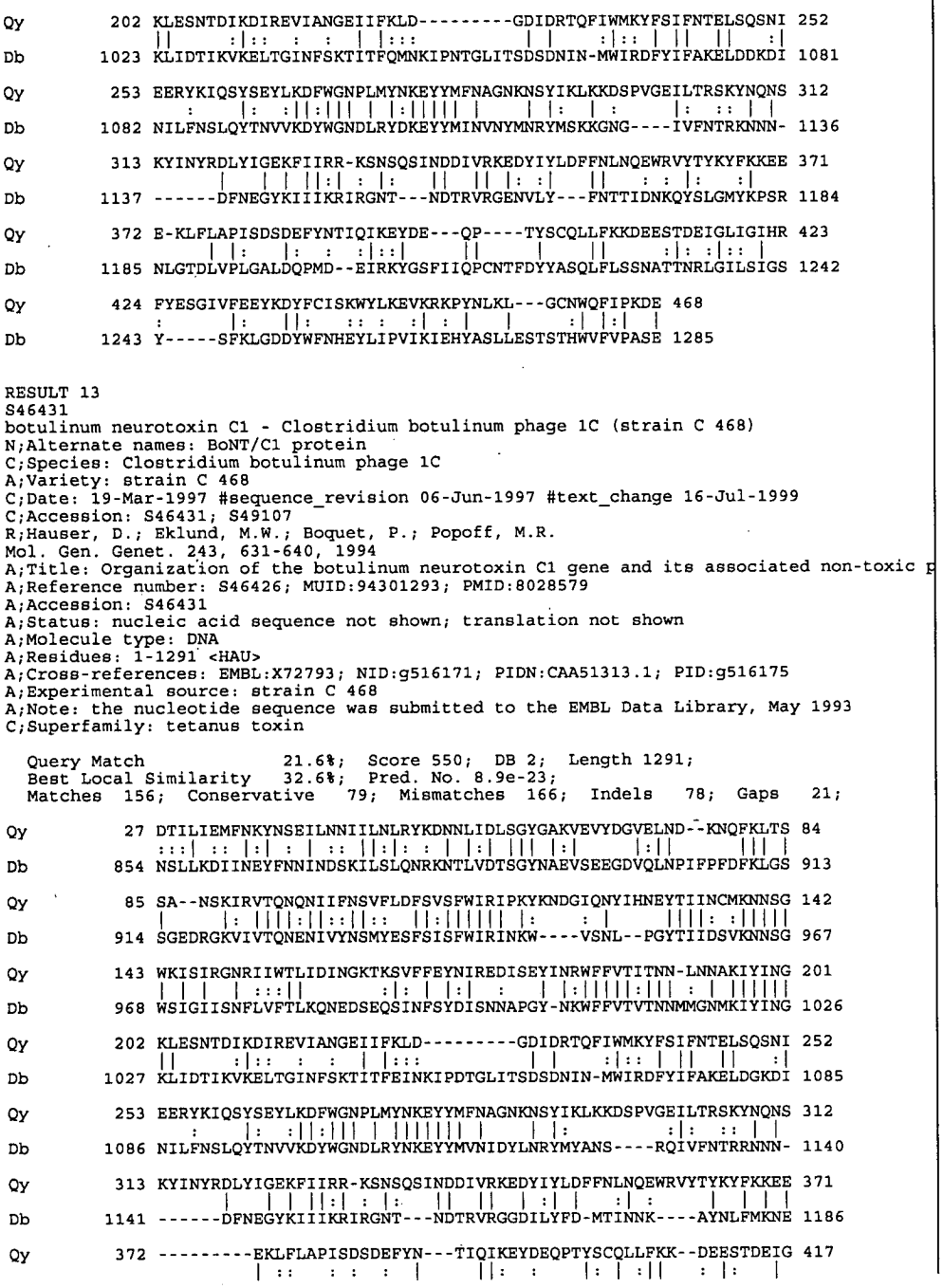

RESULT 12
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C and
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membra
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match            22.2%;  Score 564;  DB 2;  Length 1285;
  Best Local Similarity  32.7%;  Pred. No. 1.5e-23;
  Matches  153;  Conservative   88;  Mismatches  169;  Indels   58;  Gaps   19;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
              :::| :| |:| : | :: ||:|: | | |:| ||| |:| |   |:||      ||| |
Db        850 NSLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGS 909

Qy         85 SAN--SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSG 142
              | :   || ||||:||::|:::  ||:|||||| |:    : |     ||||: :|||||
Db        910 SGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSG 963

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 201
              | | |  | :::||       : : | |:| :: : | |:|||||| | : |  |||||
Db        964 WSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGY-NKWFFVTITTNMMGNMMIYING 1022
```



RESULT 13
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 21.6%; Score 550; DB 2; Length 1291;
Best Local Similarity 32.6%; Pred. No. 8.9e-23;
Matches 156; Conservative 79; Mismatches 166; Indels 78; Gaps 21;



RESULT 14
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.;
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxir
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutinir
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whol
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 21.6%; Score 550; DB 2; Length 1291;
Best Local Similarity 32.6%; Pred. No. 8.9e-23;
Matches 156; Conservative 79; Mismatches 166; Indels 78; Gaps 21;

```
Db       1086 NILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN- 1140

Qy        313 KYINYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE 371
                  |   | | ||:| : |:   ||  ||  | :| |   :| :     |  | | |
Db       1141 ------DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFD-MTINNK----AYNLFMKNE 1186

Qy        372 ---------EKLFLAPISDSDEFYN---TIQIKEYDEQPTYSCQLLFKK--DEESTDEIG 417
                    | ::   : :  :  |     ||:  :    |: | :||   : |:   |
Db       1187 TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQ-IFKSNFNGENISGIC 1245

Qy        418 LIGIHRFYESGIVFEEYKDYFCISKWY-----LKEVKRKPYNLKL---GCNWQFIPKDE 468
               || :||   |              ||     :  ||:  |   |     :| |:|  |
Db       1246 SIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 15
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             13.7%;  Score 349;  DB 2;  Length 366;
  Best Local Similarity   46.3%;  Pred. No. 2.3e-12;
  Matches   68;  Conservative   33;  Mismatches   40;  Indels    6;  Gaps    3;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  |||   :| :| ||::||::|  ||:||||: : :   | :   ::||| : |
Db        218 DKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYS 277

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: : || :||:|  : :||:|||:||||| |   :  ::||||||:|:: |||||
Db        278 SKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFN---KVNLNNEYTIIDCIRNNNSGW 334

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEY 170
              |||:  |:||||| |  |  : : | |
Db        335 KISLNYNKIIWTLQDTAGNNQKLVFNY 361


Search completed: September 10, 2004, 03:43:57
Job time : 10.3386 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 5.75902 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 2545
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2382	93.6	1290	1	BXB_CLOBO	P10844 clostridium
2	1206.5	47.4	1296	1	BXG_CLOBO	Q60393 clostridium
3	838.5	32.9	1295	1	BXA1_CLOBO	P10845 clostridium
4	784.5	30.8	1295	1	BXA2_CLOBO	Q45894 clostridium
5	740	29.1	1274	1	BXF_CLOBO	P30996 clostridium
6	713	28.0	1250	1	BXE_CLOBO	Q00496 clostridium
7	697	27.4	1250	1	BXE_CLOBU	P30995 clostridium
8	621.5	24.4	1276	1	BXD_CLOBO	P19321 clostridium
9	590	23.2	1314	1	TETX_CLOTE	P04958 clostridium
10	550	21.6	1290	1	BXC1_CLOBO	P18640 clostridium
11	238.5	9.4	1162	1	BXEN_CLOBO	P46082 clostridium
12	234.5	9.2	1162	1	BXEN_CLOBU	Q06366 clostridium
13	226	8.9	1196	1	BXCN_CLOBO	P46081 clostridium
14	158	6.2	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
15	157.5	6.2	2136	1	YCF2_MARPO	P09975 marchantia
16	152.5	6.0	339	1	RMAR_CANGA	P21358 candida gla
17	152	6.0	968	1	Y080_BUCAP	Q8ka32 buchnera ap
18	149.5	5.9	445	1	II2B_MOUSE	Q9r002 mus musculu
19	149	5.9	964	1	DPOL_CBEPV	P30319 choristoneu
20	148.5	5.8	445	1	II2A_MOUSE	P15091 mus musculu
21	147.5	5.8	544	1	FLGK_BUCAP	Q8k9k0 buchnera ap
22	146.5	5.8	519	1	YC82_ASTLO	P58145 astasia lon
23	146	5.7	360	1	MTLD_MYCPU	Q98ph2 mycoplasma
24	145	5.7	476	1	RF3_SACBA	P05512 saccharomyc
25	145	5.7	559	1	ENS2_YEAST	P12294 saccharomyc
26	144.5	5.7	641	1	PRIM_UREPA	Q9ppz6 ureaplasma
27	139.5	5.5	765	1	YG26_METJA	Q59019 methanococc
28	138.5	5.4	569	1	CYSP_PLAFA	P25805 plasmodium
29	138	5.4	544	1	OXAA_BORBU	O51398 borrelia bu
30	137	5.4	817	1	RPOD_ASTLO	P58132 astasia lon
31	136.5	5.4	2366	1	TOXB_CLODI	P18177 clostridium
32	135.5	5.3	1024	1	RPOB_PLAFA	P21421 plasmodium
33	135	5.3	464	1	YCF1_MARPO	P12223 marchantia
34	133.5	5.2	793	1	YEA8_SCHPO	O14073 schizosacch
35	133	5.2	582	1	MUTL_BUCAP	Q8k913 buchnera ap
36	132.5	5.2	1048	1	YC81_METJA	Q58677 methanococc
37	132.5	5.2	1179	1	EX5B_BUCAP	Q8k9a9 buchnera ap
38	131.5	5.2	380	1	RMAR_WILMR	P47906 williopsis
39	131.5	5.2	1018	1	YC14_METJA	Q58611 methanococc
40	131	5.1	747	1	ASNH_BACSU	P42113 bacillus su
41	131	5.1	2339	1	RPC1_PLAFA	P27625 plasmodium
42	130	5.1	971	1	Y228_BORBU	O51246 borrelia bu
43	129.5	5.1	1036	1	Y414_MYCGE	P47653 mycoplasma
44	129	5.1	715	1	UVRD_MYCPN	P75437 mycoplasma
45	129	5.1	844	1	MCEL_VACCC	P20979 vaccinia vi

ALIGNMENTS

RESULT 1
BXB_CLOBO
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence.";
RL Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.";
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;

```
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             93.6%;  Score 2382;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-126;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 904

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 1024

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 1084

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1144

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1145 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 1204

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1205 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 1264

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
               ||||||||||||||||||||||||||
Db       1265 EVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 2
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
```

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             47.4%;  Score 1206.5;  DB 1;  Length 1296;
  Best Local Similarity   50.7%;  Pred. No. 1.6e-60;
  Matches  229;  Conservative   95;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
              |||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||    |||| :
Db        850 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 909

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 144
              | || |   |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:|||
Db        910 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 969

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 203
              :||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        970 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1029

Qy        204 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 263
              : :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :
Db       1030 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1089

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 323
              |||||||| |: :||:|| | :| |||    : :||.  |: :|  :  |||::||:|
Db       1090 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 1147

Qy        324 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 381
              :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||:|
Db       1148 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 1207

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 440
                ||: :|||:| |: ||:||:| :||   |   || || :| :  | |:: | :|||
Db       1208 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 1264

Qy        441 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |:|||: :      |:|||||||| ||||||
Db       1265 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1296
```

```
RESULT 3
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
```

```
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             32.9%;  Score 838.5;  DB 1;  Length 1295;
  Best Local Similarity   39.7%;  Pred. No. 7.3e-40;
  Matches  181;  Conservative   84;  Mismatches  154;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||||||: |:||||| | :|: :   |  :  |||| :| :  :
Db        861 LLSTFTEYIKNIIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 920

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              ||| |     |::||:: :|| |||||||| |       ::|||||||||:||||||:|:
Db        921 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN---SISLNNEYTIINCMENNSGWKVSL 977

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
                 ||||| |     : | |:|:   :||:||||| ||||||| |||:||||||:|
```

```
Db        978 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1037

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              |  ::  : |:  |:|||||  |  ::||:|||::|: ||::  |:: |  || |  ||
Db       1038 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1097

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||: | |:| ||| |  : | |: :              | | ::| : |  :|
Db       1098 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSS----- 1152

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:|||  | :|::    |:|:|: |        |: |
Db       1153 --LYRGTKFIIKKYASGN--KDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1208

Qy        378 PISDSDEFYNTIQIKEYDEQP-TYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  ::|  |  |::     : : ::|| || |:|
Db       1209 EIPDVGNLSQVVVMKSKNDQGITNKCKM--NLQDNNGNDIGFIGFHQFNNIA-------- 1258

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::::|   :  |||:|:||| |:|| |
Db       1259 KLVASNWYNRQIERS--SRTLGCSWEFIPVDDGWGE 1292


RESULT 4
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             30.8%;  Score 784.5;  DB 1;  Length 1295;
  Best Local Similarity   37.9%;  Pred. No. 7.9e-37;
  Matches  173;  Conservative   89;  Mismatches  157;  Indels   37;  Gaps   10;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN--DKNQFKLTSSAN 87
              |:  | :|   |:|  ||:: || ::|||| ||||: : | |  :  |||| || :  :
Db        861 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 920

Qy         88 SKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISI 147
              | | |     |::||:: :|| ||||:|||| :   :  ::|||||||::||||||:|:
Db        921 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSL 977

Qy        148 RGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESN 206
                 ||||| |     : | |:|:   :||:|||| |||||| | :||||||:|
Db        978 NYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQ 1037

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              |  ::  : |: :|:|||||  |  ::| :|||::|: ||::  |:: |  || |  ||
Db       1038 KPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1097

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKK---------DSPVGEILTRSKYNQNSKYINY 317
              ||||| | |:| ||| |  : | |: :              | | ::| : |  ::
Db       1098 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNST----- 1152

Qy        318 RDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLA 377
              || | ||||:: ::     |:||| | :|::    |:|:|: |        |: |
Db       1153 --LYEGTKFIIKKYASGN--EDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSAL 1208

Qy        378 PISDSDEFYNTIQIKEYDEQPTYS-CQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD 436
              | |       : :|  |:|  : |::     : : ::|| || | :      :
Db       1209 EIPDVGNLSQVVVMKSKDDQGIRNKCKM--NLQDNNGNDIGFIGFHLY--------DNIA 1258

Qy        437 YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                | || ::| :   :   ||:|:||| |:|| |
```

```
Db       1259 KLVASNWYNRQVGKA--SRTFGCSWEFIPVDDGWGE 1292


RESULT 5
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             29.1%;  Score 740;  DB 1;  Length 1274;
  Best Local Similarity   37.0%;  Pred. No. 2.4e-34;
  Matches  171;  Conservative   82;  Mismatches  155;  Indels   54;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  ||:   :| :: ||::||::|  ||:||||: : :   | :   ::||| : :
Db        851 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 910

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGW 143
              |  |:: : || :||:|| : :||:|||:||||:      |    |||||||| |||||
Db        911 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGW 967

Qy        144 KISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYI 199
              |||:|  |   ||||| | :|  ::: | |     || |||:| ||||||| | |::|||
Db        968 KISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYI 1027

Qy        200 NGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQ 259
              || |     | :: ::  :  |:||: |  | | :: ::|| :||||| :: ||  |  :
Db       1028 NGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVVFNTELDKTEIETLYSNE 1086

Qy        260 SYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRD 319
              ||:|||  |:|||:||:||    |: || |  :| :  |  :    : | ::||:
Db       1087 PDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK- 1143

Qy        320 LYIGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLF--- 375
              || | : |||:       | |: ||| |  |::  :   |:|:|     | |:||:
Db       1144 LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIRTS 1201

Qy        376 -----LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV 430
                  |  |   |  |             :| : |: :  |   |||:| |
Db       1202 NLNDSLGQIIVMDSIGN------------NCTMNFQNNNGS--NIGLLGFH-------- 1238

Qy        431 FEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                :   | ||   ::|       || |  | |: || |
Db       1239 ----SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 1274


RESULT 6
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
```

```
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             28.0%;  Score 713;  DB 1;  Length 1250;
  Best Local Similarity   35.5%;  Pred. No. 7.8e-33;
  Matches  160;  Conservative   91;  Mismatches  162;  Indels   38;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:|||::  :| ||| : : :   |     :|||| : :
Db        832 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 891

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
                |:: ::||  ||::: : :||:|||:||| | |  :   ::||||||||: |||||
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 949

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
              |:|:  | ||||  |  |  : : | |     ||:|||:| ||||||: | ::|:||||
Db        950 KVSLNHNEIIWTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
              |      | ::  :  :  |:||:  :   |::| ::||:||: || :: |:  |  :  :
Db       1010 LIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1068

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
              ||||||| |:|:|||||: |    |::|  :|||      | |  |  |  :    ||
Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS------TLSINNIRSTILLANRLYS 1122

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTYKYFKKEEEKLFLAPISD 381
              | |  |:| :|| | ||::||| | :|::|  :    : :|       :|: :    ||
```

```
Db       1123 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTI---KISS 1178

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS 441
                  |   :| : :         :| : || :  : : |||:|               |    |
Db       1179 SGNRFNQVVVM----NSVGNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1220

Qy        442 KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                ||   :  : :    || | || :: || |
Db       1221 TWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1249


RESULT 7
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             27.4%;  Score 697;  DB 1;  Length 1250;
  Best Local Similarity   35.3%;  Pred. No. 6.2e-32;
  Matches  159;  Conservative   91;  Mismatches  163;  Indels   38;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
                | |||  |||:   | :: :||:|||::  :| ||| : : :   |     :|||| : :
Db        832 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 891

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
                 |:: ::||  ||::: : :||:|||:||| | |  :   ::|||||||||: |||||
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 949

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
                |:|:  | ||||| | :|  : : | |     ||:|||:| ||||||: | ::|:||||
Db        950 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
                |     | ::  :  :  |:||:  :   |::| ::||:||: || :: |:  |  :  :
Db       1010 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1068

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
                |||||| |:|:||||: |    |::|  : ||      | |  |  |  :    ||
Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS------TLSINNIRSTILLANRLYS 1122

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTYKYFKKEEEKLFLAPISD 381
                | |  |:| :|| | ||::||| | :|::|  :      :|       :|: :    ||
Db       1123 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTI---KISS 1178

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS 441
                  |   :| : :         :| : || :  : : |||:|               |    |
Db       1179 SGNRFNQVVVM----NSVGNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1220

Qy        442 KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                ||   ::    :   |  | || :: |+| |
Db       1221 TWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249


RESULT 8
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
```

```
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             24.4%;  Score 621.5;  DB 1;  Length 1276;
  Best Local Similarity   33.9%;  Pred. No. 1.1e-27;
  Matches  160;  Conservative   78;  Mismatches  163;  Indels   71;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
              :::| :: |:| : | :: ||:|: | | |:| ||| |:| | | |:||    | |||:|
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 144
              | : || |  | ||:::::: : ||||||:| |   :      |||||||| :: |||||
Db        910 SGD-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWK 963

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 203
              : ||   | | | |:| | ||: |:|:     : | |:|||||||| :   |:||||:|
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        204 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 263
              : :  |:|: ||  :  |:| :| :||  | :|:: |:||: |||  :|   |: |
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 323
              :||:||||| :: ||||: |    : ||     :|   :|   :|   ||       || |
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVQYPDRSK------LYTG 1132

Qy        324 EKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSD 383
                |:  |:    :   :   | | |    :::: :                 | |:|
Db       1133 NPITIKSVSDKNPYSR--ILNGDNIILHMLYNSRKYMI----------------IRDTD 1173

Qy        384 EFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV-----------F 431
              | | |  |  :   |: :|          :|  ||  |    ||             |
Db       1174 TIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSF 1224

Qy        432 EE--------YKDYFCISKWYLKEVKRKPYNLKL---GCNWQFIPKDEGWTE 472
              |         || :     |     |  ||       |:|| :| || |
Db       1225 RENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 9
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
```

```
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0     0
FT   CHAIN         1   456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457  1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232   232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233   233       BY SIMILARITY.
FT   METAL       236   236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226   246       POTENTIAL.
FT   TRANSMEM    669   689       POTENTIAL.
FT   DISULFID    438   466       INTERCHAIN.
FT   DISULFID   1076  1092
FT   HELIX       876   882
FT   TURN        883   883
FT   STRAND      884   891
FT   TURN        892   893
FT   STRAND      894   897
FT   STRAND      904   907
FT   TURN        909   910
FT   STRAND      912   915
```

```
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190

  Query Match           23.2%;  Score 590;  DB 1;  Length 1314;
  Best Local Similarity  31.7%;  Pred. No. 6.6e-26;
  Matches  140;  Conservative   89;  Mismatches  184;  Indels   28;  Gaps   12;

Qy         46 ILNLRYKDNNLIDLSGYGAKVEVYDGVEL----NDKNQFKLTSSANSKIRVTQNQNIIFN 101
              ||||   :: : |:||: : |  |   :|    | |    | :: :|:: | :  :| :|
Db        885 ILNLDINNDIISDISGFNSSVITYPDAQLVPGINGK-AIHLVNNESSEVIVHKAMDIEYN 943

Qy        102 SVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNN-----SGWKISIRGNRIIWTL 156
              :| :|:||||:|:||     :: |  |||:||: || :     ||| :|::|| :|||
Db        944 DMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTL 1003

Qy        157 IDINGKTKSVFFEYNIREDISEYI-NRWFFVTITNN-LNNAKIYINGKLESNTDIKDIRE 214
              |  |: : : |  :: :  : |: |:| |:||||: |::| :|||| |  : :|  :
Db       1004 KDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGA 1062

Qy        215 VIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLM 274
              :  :  |  |||   :  |:: :  | ||   |:   ||: |       :|:|||||||
Db       1063 IREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLR 1122

Qy        275 YNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNS 334
              |: |||:    : :  ::||  :    :     |      | || || | ||||:| : :
Db       1123 YDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPN 1182

Qy        335 QSINDDIVRKEDYIYLDFFNLNQEWRVYTY----KYFKKEEEKLFLAPISDSDEFYNTIQ 390
              | |  |:  |:| | : : |    :  |       |   :  | :   :     | ::
Db       1183 NEI-DSFVKSGDFIKL-YVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKME 1240

Qy        391 IKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEVKR 450
              :  :  ||| ||    |: ::  :||:| |       |  :  :|    | ||   :|
Db       1241 AVKLRDLKTYSVQLKLYDDKNAS--LGLVGTH---NGQIGNDPNRDILIASNWYFNHLKD 1295

Qy        451 KPYNLKLGCNWQFIPKDEGWT 471
              |    |||:| |:| |||||
Db       1296 K----ILGCDWYFVPTDEGWT 1312


RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
```

```
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             21.6%;  Score 550;  DB 1;  Length 1290;
  Best Local Similarity   32.6%;  Pred. No. 1.1e-23;
  Matches  156;  Conservative   79;  Mismatches  166;  Indels   78;  Gaps   21;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
              :::|:: :|:| : | :: ||:|: : | |:| ||| |:|     |:||     ||| |
Db        853 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 912

Qy         85 SA--NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSG 142
              |     |: ||||:||::||::  ||:|||||| |:    : |     ||||: :|||||
Db        913 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSG 966

Qy        143 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 201
              | | |  | :::||       :|: | |:|  :   | |:|||||:||| : | ||||||
Db        967 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1025

Qy        202 KLESNTDIKDIREVIANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNI 252
              ||     :|:: :  :  | |:::          | |    :|:: | ||  ||   :|
Db       1026 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDI 1084

Qy        253 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNS 312
                 :   |: :||:||| | ||||||| |    | |:          :|:  :: | |
Db       1085 NILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN- 1139

Qy        313 KYINYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE 371
                  |   | | ||:| : |:   ||  ||  | :| |   :| :     |  | | |
Db       1140 ------DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFD-MTINNK----AYNLFMKNE 1185

Qy        372 ---------EKLFLAPISDSDEFYN---TIQIKEYDEQPTYSCQLLFKK--DEESTDEIG 417
                     | ::   : :  :  |     ||:  :    |: | :||   : |:   |
Db       1186 TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQ-IFKSNFNGENISGIC 1244

Qy        418 LIGIHRFYESGIVFEEYKDYFCISKWY-----LKEVKRKPYNLKL---GCNWQFIPKDE 468
                 || :||   |              ||      :  ||:  |   |     :| |:|  |
Db       1245 SIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1290


RESULT 11
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;

  Query Match              9.4%;  Score 238.5;  DB 1;  Length 1162;
  Best Local Similarity   23.5%;  Pred. No. 3e-06;
  Matches   91;  Conservative   73;  Mismatches  133;  Indels   91;  Gaps   18;

Qy         33 MFNKYNSEILNN-------IILNLRYKDNNLI-DLSGYGAKVEVYDGVELNDKNQFKLTS 84
              :|| |   ::           :   : :|||:| | ||     ||    : |       :
Db        784 LFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGKNTLVEYPKDIGL-------VYG 836

Qy         85 SANSKIRVT-QNQNIIF------NSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM 137
              |: | :|  |||| |       | :  :||: ||:|  |       || | ::  :|
Db        837 INNNAIHLTGANQNIKFTNDYFENGLTNNFSIYFWLRNLK------QNTIKSK--LIGSK 888

Qy        138 KNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR-WFFVTITNNLNNAK 196
              ::| ||:|    : ::: :|| ||  |:::          :|   |: | :: |  ::|  |
Db        889 EDNCGWEIYFENDGLVFNIIDSNGNEKNIY--------LSNISNKSWHYIVI--SINRLK 938

Qy        197 ----IYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 252
                  |:|:  | :| |||:|  : ::  |     | :      ::::  |: |  :: : |
Db        939 DQLLIFIDNILVANEDIKEILNIYSSDIISLLSDNN-----NVYIEGLSVLNKTINSNEI 993

Qy        253 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFN----------AGNKNSYIKLKKDSPVGE 301
                  |     : |:::|     | ||: | :||                | | |: :  :: :
Db        994 LTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEIAINKIEQNNNIYLSINNENNLNF 1053

Qy        302 ILTRSKY---NQNSKYINYRDLYI------GEKF----------IIRRKSNSQ--SIND 339
                 : |    | | :|:   |  |        ||:           ::  |:|:|   ||:
```

```
Db       1054 KPLKFKLLNTNPNKQYVQKWDEVIFSVLDGTEKYLDISTTNNRIQLVDNKNNAQIFIINN 1113

Qy        340 DI--------VRKEDYIYLDFFNLNQEW 359
              ||              :||   | : |
Db       1114 DIFISNCLTLTYNNVNVYLSIKNQDYNW 1141


RESULT 12
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match             9.2%;  Score 234.5;  DB 1;  Length 1162;
  Best Local Similarity   23.2%;  Pred. No. 5.1e-06;
  Matches   90;  Conservative   73;  Mismatches  134;  Indels   91;  Gaps   18;

Qy         33 MFNKYNSEILNN-------IILNLRYKDNNLI-DLSGYGAKVEVYDGVELNDKNQFKLTS 84
              :|| |   ::          :   : :|||:| | ||    ||    : |       :
Db        784 LFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGKNTLVEYPKDIGL-------VYG 836

Qy         85 SANSKIRVT-QNQNIIF------NSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM 137
                |: | :|  |||| |      | :  :||: ||:|      :  || | ::  :|
Db        837 INNNAIHLTGANQNIKFTNDYFENGLTNNFSIYFWLR------NLNQNTIKSK--LIGSK 888

Qy        138 KNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR-WFFVTITNNLNNAK 196
              ::| ||:|    | ::: :|| ||  |:::        :|   |: | :: | ::|  |
Db        889 EDNCGWEIYFENNGLVFNIIDSNGNEKNIY--------LSNISNKSWHYIVI--SINRLK 938

Qy        197 ----IYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 252
                  |:|:  | :| |||:|  : ::  |    | :      ::::  |: |  :: : |
Db        939 DQLLIFIDNILVANEDIKEILNIYSSDIISLLSDNN-----NVYIEGLSVLNKTINSNEI 993

Qy        253 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFN-----------AGNKNSYIKLKKDSPVGE 301
                 |     : |:::|    | ||: | :||             | | |:    :: :
Db        994 LTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEIAINKIEQNNNIYLSNNNENSLNF 1053

Qy        302 ILTRSKY---NQNSKYI--------------NYRDLYIGEKFI--IRRKSNSQS--IND 339
                : |    | | :|:                | |: |    |  :  |:|:::  ||:
Db       1054 KPLKFKLLNTNPNKQYVQKWDEVIFSVLDGTEKYLDISIDNNRIQLVDNKNNAKTFIINN 1113

Qy        340 DI--------VRKEDYIYLDFFNLNQEW 359
              ||              :||   | : |
Db       1114 DIFISNCLTLTYNNVNVYLSIKNQDYNW 1141


RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             8.9%;  Score 226;  DB 1;  Length 1196;
  Best Local Similarity   20.4%;  Pred. No. 1.6e-05;
  Matches   85;  Conservative   86;  Mismatches  153;  Indels   92;  Gaps   19;

Qy         24 SMADTILIEMFNKYNSEILNNIILNLRYKDNNLI-DLSGYGAKVEVYDGVEL----NDKN 78
              |  : :: |: : |       ::   :  ::|:| |:||    ::  : :||    | ::
Db        822 SQVEQVMKEILSPYQL-----LLFASKGPNSNIIEDISGKNTLIQYTESIELVYGVNGES 876

Qy         79 QFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK 138
              :    | |  |:   :     | :  :|:: ||:|    |||     | |:
Db        877 LY--LKSPNETIKF--SNKFFTNGLTNNFTICFWLRFTG-KNDDKTRLIGNKV------- 924

Qy        139 NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYIN-RWFFVTITNN--LNNA 195
              || ||:|    | ::: :|| ||  :||:        :|  ||  |::::|: :   :
Db        925 NNCGWEIYFEDNGLVFEIIDSNGNQESVY--------LSNIINDNWYYISISVDRLKDQL 976

Qy        196 KIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEER 255
              |:|| |  :|  |  |  : :   |       :::   |:::  |: :  ::   :
Db        977 LIFINDKNVANVSIDQILSIYSTNIISL-----VNKNNSIYVEELSVLDNPITSEEVIRN 1031

Qy        256 YKIQSYSEYLKDFWGNPLMYNKEYYMFN-----------AGNKNSYIKLKKDSPVGEILT 304
```

```
Db       1032 YFSYLDNSYIRDSSKSLLEYNKNYQLYNYVFPETSLYEVNDNNKSYLSLKNTDGINISSV 1091

Qy        305 RSK---YNQNSKYINYRDLYI------GEKF-----------IIRRKSNSQ--SINDDIV 342

Db       1092 KFKLINIDESKVYVQKWDECIICVLDGTEKYLDISPENNRIQLVSSKDNAKKITVNTDLF 1151

Qy        343 RKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQP 398

Db       1152 RPD--------------CITFSY----NDKYFSLSLRDGD--YNWMICNDNNKVP 1186


RESULT 14
EX5B_BUCBP
ID   EX5B_BUCBP     STANDARD;      PRT;  1180 AA.
AC   Q89AB3;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN   RECB OR BBP404.
OS   Buchnera aphidicola (subsp. Baizongia pistaciae).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=135842;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426901; PubMed=12522265;
RA   Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA   Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA   Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT   "Reductive genome evolution in Buchnera aphidicola.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC   -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
CC       UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC       STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION. ALL OF
CC       THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC       ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC       phosphooligonucleotides.
CC   -!- SUBUNIT: Consist of three subunits; recB, recC and recD (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the helicase family. UvrD subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE014017; AAO27116.1; -.
DR   InterPro; IPR000212; UvrD-helicase.
DR   Pfam; PF00580; UvrD-helicase; 1.
KW   Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW   DNA repair; Complete proteome.
FT   NP_BIND      24     31       ATP (POTENTIAL).
SQ   SEQUENCE   1180 AA;  139435 MW;  55E1C0F9F389FAE9 CRC64;

  Query Match             6.2%;  Score 158;  DB 1;  Length 1180;
  Best Local Similarity   22.7%;  Pred. No. 0.1;
  Matches   93;  Conservative   63;  Mismatches  128;  Indels  126;  Gaps   21;

Qy         30 LIEMFNKYNSEILNNIILNLRYKDNNLIDL---------SGYGAKVEVYDGVELNDKNQF 80

Db        382 LIDEFQDTNNKQYN--IFKKIYKKENLLILISDPKQAIYSFSGADITSYLQAKSNIKNQY 439

Qy         81 KLTSSANSKIRVTQNQNIIFNSV---FLDFSVSFW----------------------IRI 115

Db        440 FLDTNWRSSPKIINSINLLFSRLKNPFLTKNITFYPVKSSRINKTTKFEVNGKNQPALRF 499

Qy        116 PKYKNDGIQNYIHNEYTIINCMKNNSGWKIS-IRGNRIIWTLIDINGKTKSVFFEYNIRE 174

Db        500 LLNKNKNISINEYKEWISITTAKYISYWISSGIKGN----ATITISNKVRTVNF-----S 550

Qy        175 DISEYINRWFFVTITNN-----------LNNAKIYINGKLESNTDIKDIREVIANGEII 222

Db        551 DIA--------IIVRNNLEAKTIQLELTKLNIQSLY----LTSKNNIFQSKEIL---EIL 595

Qy        223 FKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKI----------QSYSEYLKDFWGN 271

Db        596 WILQAILNPNIERLLK--RAMSTNIMSKNTKEIISIPNNHSYWIKLSQEFNQYL-IFWNN 652

Qy        272 PLMYNKEYYM------FNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 325

Db        653 ---YGILYVIQQLIINYNISNNNNLL-------------HNFSPNIKNI----LYIGE- 690

Qy        326 FIIRRKSNSQSINDDIVRKEDYIYLDFF--NLNQEWRVYTYKYFKKEEEK 373

Db        691 -LLEEKS--------ITIKNKFLLINWLKKNITQDFYLTKPKYIKPNYEK 731


RESULT 15
YCF2_MARPO
ID   YCF2_MARPO     STANDARD;      PRT;  2136 AA.
AC   P09975;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN   YCF2.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC   Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
CC   -!- SIMILARITY: Belongs to the ycf2 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04465; CAA28078.1; -.
DR   PIR; S01591; A05037.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
KW   Chloroplast; Hypothetical protein.
SQ   SEQUENCE   2136 AA;  259911 MW;  5BD170C9CCF61197 CRC64;

  Query Match             6.2%;  Score 157.5;  DB 1;  Length 2136;
  Best Local Similarity   19.0%;  Pred. No. 0.21;
  Matches  100;  Conservative   96;  Mismatches  191;  Indels  139;  Gaps   24;

Qy         37 YNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQ 96

Db        149 YKNSLLNLIDFRSILKKKEILNLHWW--KFLVLEQIQSN----WKISEESLSELKIVLEQ 202

Qy         97 NIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM----------KNNSGWKIS 146
```

```
Db        203 K---NIDELKHFFEFYINQKIYPNNNWEYYFYSIF--INQLKIDIKNSKYNKNSIGFEVF 257

Qy        147 IRGNRIIWTLIDINGK--TKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLE 204

Db        258 LAFCEKLLFEVEFLSKPNNNNLQMKLNCLENFS-------FLDIFCILNKKLPWVNKKIF 310

Qy        205 SN-TDIKDIREVIANGEIIFKLDGDIDRTQFI----WMKY------FSIFNTELSQS--- 250

Db        311 KNLQNFNESDKKLIESFFLLKIKGNLYFKNYIEFVTWQSYKKDCLDFNKFN-ELNNSEIY 369

Qy        251 -NIEER-----YKIQSYSEY--------LKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKD 296

Db        370 IKIEELFSDYIYKFSKYILYEGKKSKTIIKQSFNNNIYYKKLNSIFNF---NTIFYFDSN 426

Qy        297 SPVGEILTRSKYNQNSKYINYRDLY--IGEKFII-RRKSNSQSINDDIVRKED------- 346

Db        427 NLLFDWLKKNYYINNKPFLKSFLIYSSISNQFILFFKQKNSKSFNKNLVKKNSKDVITNV 486

Qy        347 ----------------YIYLDFFNLNQEWRVYTYKYF------KKEEEKLFLAPISDSD 383

Db        487 FSKENKIEINNFSKSIYYAFFEILSINEIDNKFVINKISLKNINKKKQKRFYLNKIKSSD 546

Qy        384 --EFYNTIQIKEYDEQ-----------PTYSCQ-----------LLFKKDEESTDEIGLI 419

Db        547 NFRFINLWKIKNYSSQQFVSNNSFLLNPAFEILQQNYYLKKKNILFFKK---------- 595

Qy        420 GIHRFYESGIVFEEYK--------DYFCISKWYLKEVKRKPYNLKL 457

Db        596 -LNEVFSNFFYFQYYKCKKLNIFLKFASLEKILKKRNKKFTISIKL 640
```

Search completed: September 10, 2004, 03:35:18
Job time : 8.75902 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 21.5963 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-46
Perfect score: 2545
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....YNLKLGCNWQFIPKDEGWTE 472

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2267	89.1	1291	2	Q93G71	Q93g71 clostridium
2	2265	89.0	1291	2	Q933K0	Q933k0 clostridium
3	2259	88.8	1291	2	Q9ZAJ8	Q9zaj8 clostridium
4	2188	86.0	1291	2	Q8GR96	Q8gr96 clostridium
5	2133	83.8	441	2	Q9X708	Q9x708 clostridium
6	2125	83.5	1291	2	Q08077	Q08077 clostridium
7	793	31.2	1268	2	Q45851	Q45851 clostridium
8	751	29.5	1278	2	Q57236	Q57236 clostridium
9	742.5	29.2	1280	2	Q9ZAJ5	Q9zaj5 clostridium
10	738	29.0	361	2	Q45848	Q45848 clostridium
11	738	29.0	361	2	Q45846	Q45846 clostridium
12	704	27.7	1251	2	Q9K395	Q9k395 clostridium
13	695.5	27.3	1252	2	Q8KZM3	Q8kzm3 clostridium
14	695.5	27.3	1255	2	Q9FAR6	Q9far6 clostridium
15	616.5	24.2	1275	12	Q9QTG7	Q9qtg7 clostridium
16	612.5	24.1	1280	2	Q45849	Q45849 clostridium
17	612.5	24.1	1280	2	Q9LBS7	Q9lbs7 clostridium
18	612.5	24.1	1280	2	Q841S3	Q841s3 clostridium
19	588	23.1	451	2	Q9LA13	Q9la13 clostridium
20	564	22.2	1285	2	Q45967	Q45967 clostridium
21	563	22.1	1285	2	Q9LBR1	Q9lbr1 clostridium
22	561	22.0	1310	2	Q93N27	Q93n27 clostridium
23	550	21.6	1291	2	Q93HT3	Q93ht3 clostridium
24	374	14.7	78	2	Q7WUH9	Q7wuh9 clostridium
25	370	14.5	78	2	Q7WRW0	Q7wrw0 clostridium
26	327.5	12.9	367	2	Q45861	Q45861 clostridium
27	327.5	12.9	367	2	Q45862	Q45862 clostridium
28	268.5	10.6	1198	2	O06018	O06018 clostridium
29	246.5	9.7	1197	2	Q45888	Q45888 clostridium
30	238	9.4	1160	2	O87710	O87710 clostridium
31	237	9.3	1196	2	Q9LBS8	Q9lbs8 clostridium
32	235	9.2	1197	2	P71117	P71117 clostridium
33	235	9.2	1197	2	O33871	O33871 clostridium
34	234.5	9.2	1197	2	Q9ZAJ9	Q9zaj9 clostridium
35	234	9.2	1196	2	Q9LBR2	Q9lbr2 clostridium
36	233	9.2	1198	2	O69277	O69277 clostridium
37	232	9.1	1198	2	Q45893	Q45893 clostridium
38	226	8.9	1196	2	Q45916	Q45916 clostridium
39	226	8.9	1196	2	Q93HT4	Q93ht4 clostridium
40	226	8.9	1196	9	Q9ZX77	Q9zx77 clostridium
41	226	8.9	1196	9	Q38197	Q38197 clostridium
42	225	8.8	1196	2	Q53550	Q53550 clostridium
43	224.5	8.8	1161	2	Q45891	Q45891 clostridium
44	223	8.8	1165	2	Q45844	Q45844 clostridium
45	223	8.8	1165	2	Q45887	Q45887 clostridium

ALIGNMENTS

RESULT 1
Q93G71
ID Q93G71 PRELIMINARY; PRT; 1291 AA.
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF295926; AAK97132.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 89.1%; Score 2267; DB 2; Length 1291;
Best Local Similarity 94.8%; Pred. No. 7.6e-109;
Matches 423; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              :|||||:||||||:|||||||||||:|| ||||||||||||||||:|||||||||||||
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              |||||| |||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              |||| |||||||||| |||||||:|:||||||||||||||| :||||||||||||:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNSDNAKIYINGKLESH 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||:|||||||| |||||||:|||||||||||||||||||| ||||||||||
Db       1026 IDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYIYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||||| ||||||||||||||||||
Db       1266 EVKRKPYNSKLGCNWQFIPKDEGWTE 1291
```

RESULT 2
Q933K0
ID Q933K0 PRELIMINARY; PRT; 1291 AA.
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=593, and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;

Query Match 89.0%; Score 2265; DB 2; Length 1291;
Best Local Similarity 94.8%; Pred. No. 9.6e-109;
Matches 423; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

```
Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              :|||||:||||||:|||||||||||:|| ||||||||||||||||:|||||||||||||
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              |||| |||||||||| |||||||:|:||||||||||||||| :||||||||||||:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNSDNAKIYINGKLESH 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||:|||||||| |||||||:|||||||||||||||||||| ||||||||||
Db       1026 IDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 446
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||||| ||||||||||||||||||
Db       1266 EVKRKPYNSKLGCNWQFIPKDEGWTE 1291
```

RESULT 3
Q9ZAJ8
ID Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match           88.8%;  Score 2259;  DB 2;  Length 1291;
  Best Local Similarity   94.6%;  Pred. No. 2e-108;
  Matches  422;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              :|||||:||||||:||||||||||:|| ||||||||||||||||||:|||||||||||||
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSA 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              |||||| |||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              |||| |||||||||||| |||||||:|:|||||||||||||||||| :||||||||||:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNSDNAKIYINGKLESH 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
               ||:|||||||| |||||||:|||||||||||||||||||||||| ||||||||||
Db       1026 IDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLFLAPISDSDEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1206 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFKEYKDYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||||| ||||||||||||||||
Db       1266 EVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 4
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match           86.0%;  Score 2188;  DB 2;  Length 1291;
  Best Local Similarity   91.7%;  Pred. No. 8.9e-105;
  Matches  409;  Conservative   18;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              :|||||:|||||||||||||||:|||||||||| ||||||||||||||||||
Db        846 NTILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELNDKNQFKLTSST 905

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||:|||||||||||:|||||||||||||||||||||||||||||||||:|||||||||
Db        906 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKIS 965

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              |||||||||| ||||||||||||:|||||:|||||||||||| :||||||||||||
Db        966 IRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNSDNAKIYINGKLESN 1025

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              |||||||||||||||||||||||||||||||||||||||||||:| ||||||||||
Db       1026 IDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYLK 1085

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              |||||||||||||||||||||||||||||| |||||||||||||||||||:|||||||
Db       1086 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEKF 1145

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              |||||||||||||||||||||||||||||| |:||||| || ||:||:||||| | ||:|||
Db       1146 IIRRKSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLFLANIYDSNEFY 1205

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              |||||||||||||||||||||||||||||||||||||||||||| ::||:||||||||||
Db       1206 KTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFCISKWYLK 1265

Qy        447 EVKRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||||| |||||||||||||| |
Db       1266 EVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 5
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;
```

```
  Query Match             83.8%;  Score 2133;  DB 2;  Length 441;
  Best Local Similarity   91.3%;  Pred. No. 1.9e-102;
  Matches  400;  Conservative   19;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSA 86
              :|||||:||||||:|||||||||||:|||||||||| ||||||||||||||||||||
Db          4 NTILIEIFNKYNSKILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELNDKNQFKLTSST 63

Qy         87 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 146
              ||:|||||||||||:|||||||||||||||||||||||||||||:|||||||||
Db         64 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKIS 123

Qy        147 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESN 206
              |||||||||| ||||||||||||:||||||:|||||||||||| :|||||||||||||
Db        124 IRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNSDNAKIYINGKLESN 183

Qy        207 TDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLK 266
              ||||| |||||||||||||||||||||||||||||||||||||||:| ||||||||||||
Db        184 IDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYLK 243

Qy        267 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKF 326
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||:|||||||
Db        244 DFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEKF 303

Qy        327 IIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFY 386
              |||||||||||||||||||||||||||||| |:||||| || ||:||:|| || | ||:|||
Db        304 IIRRKSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLVLANIYDSNEFY 363

Qy        387 NTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLK 446
              |||||||||||||||||||||||||||||||||||||||||||||||| ::||:||||||||||
Db        364 KTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFCISKWYLK 423

Qy        447 EVKRKPYNLKLGCNWQFI 464
              ||||||||  ||||||||
Db        424 EVKRKPYNPNLGCNWQFI 441
```

```
RESULT 6
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;
```

```
  Query Match             83.5%;  Score 2125;  DB 2;  Length 1291;
  Best Local Similarity   88.7%;  Pred. No. 1.6e-101;
  Matches  394;  Conservative   28;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANS 88
              |||::|||||||||||||||:||||||||||||||||||:|||||||||||||:|
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 907

Qy         89 KIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||:|||||||||||:|| ||||||||||||||||||||||||||||
Db        908 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 967

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTD 208
              |||||||||||||||||||||||||||||||||||||||||||||:|||||||| |||| |
Db        968 GNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLDNAKIYINGTLESNMD 1027

Qy        209 IKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDF 268
              |||| ||| |||| ||||||:|||||||||||||||:|:||||:| ||||||||||||
Db       1028 IKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDF 1087

Qy        269 WGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFII 328
              |||||||||||||||||||||||| ||| ||||| |||||||| |||||:|||||||||
Db       1088 WGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFII 1147

Qy        329 RRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEFYNT 388
              ||:|||||||||||||:||    ::||||| |||||::||||||: ||||:||| |
Db       1148 RRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISDSNEFYKT 1207

Qy        389 IQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWYLKEV 448
              |:||||||||:|||||||||||||:|||||||||||:: ::|||||||||||||||
Db       1208 IEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEV 1267

Qy        449 KRKPYNLKLGCNWQFIPKDEGWTE 472
              |||||   ||||||||||||||||
Db       1268 KRKPYKSNLGCNWQFIPKDEGWTE 1291
```

```
RESULT 7
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
```

```
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match           31.2%;  Score 793;  DB 2;  Length 1268;
  Best Local Similarity  39.2%;  Pred. No. 5.8e-33;
  Matches  177;  Conservative   93;  Mismatches  150;  Indels   32;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
Db        843 DKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYS 902

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
Db        903 SRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN--LKN-LNNEYTIINCMRNNNSGW 959

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
Db        960 KISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGN 1019

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
Db       1020 LTDQKSILNLGNIHVDDNILFKIVGCND-TRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEI-LTRSKYNQNSKYINYRDLY 321
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPN--MSVTKNSDILNINRQRGIYSKTNIFSNAR-LY 1135

Qy        322 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPIS 380
Db       1136 TGVEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRIS 1195

Qy        381 DSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCI 440
Db       1196 NSN--YNSNQMIIMDSIGD-NCTMNFKTN--NGNDIGLLGFH-----------LNNLVA 1238

Qy        441 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
Db       1239 SSWYYKNIRNNTRN--NGCFWSFISKEHGWQE 1268


RESULT 8
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           29.5%;  Score 751;  DB 2;  Length 1278;
  Best Local Similarity  35.6%;  Pred. No. 8.4e-31;
  Matches  166;  Conservative   88;  Mismatches  152;  Indels   60;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
Db        852 DKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYS 911

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
Db        912 SKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFN---KVNLNNEYTIIDCIRNNNSGW 968

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
Db        969 KISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGN 1028

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
Db       1029 LIDEKSISNLGDIHVSDNILFKIVGCND-TRYVGIRYFKVFDTELGKTEIETLYSDEPDP 1087

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRD--- 319
Db       1088 SILKDFWGNYLLYNKRYYLLNL--------LRTDKSI---------TQNSNFLNINQQRG 1130

Qy        320 ------------LYIGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKY 366
Db       1131 VYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADIS 1190

Qy        367 FKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYE 426
Db       1191 IAKPEKIIKLIRTSNSNNSLGQIIVM---DSIGNNCTMNFQNN--NGGNIGLLGFH---- 1241

Qy        427 SGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
Db       1242 --------SNNLVASSWYYNNIRKN--TSSNGCFWSFISKEHGWQE 1277
```

```
RESULT 9
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             29.2%;  Score 742.5;  DB 2;  Length 1280;
  Best Local Similarity   37.1%;  Pred. No. 2.3e-30;
  Matches  171;  Conservative   86;  Mismatches  157;  Indels   47;  Gaps   15;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  ||:   :| :: ||::||::|  ||:||||: : :   | :   ::||| : |
Db        852 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYS 911

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGW 143
                |:: : || :||:|| : :||:|||: |||:     :|    ||||||||  |||||
Db        912 GRLSEVNIAQNNDIIYNSRYQNFSISFWVTIPKHYRPMNRN---REYTIINCMGNNNSGW 968

Qy        144 KISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYI 199
              |||:|  |   ||||| | :|  : : | |      ||:|||:| ||||||| | |::|||
Db        969 KISLRTIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGNSRIYI 1028

Qy        200 NGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQ 259
              || |      | :: ::  :  |:||: |  | | :: ::|| :||||| :: ||  |  :
Db       1029 NGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNE 1087

Qy        260 SYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQN-------S 312
                 |||:||| |:|||:||:||    |: ||   ::| :  |      ||          |
Db       1088 PDPSILKDYWGNYLLYNKKYYLFNLLRKDKYI--TRNSGILNI------NQQRGVTGGIS 1139

Qy        313 KYINYRDLYIGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE 371
              ::||: || | : |||: :     | |: ||| |  |::  :   |:|:|       | |
Db       1140 VFLNYK-LYEGVEVIIRKNAPIDISNTDNFVRKNDLAYINVVDHGVEYRLYADISITKSE 1198

Qy        372 EKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVF 431
              : : |   |: ::    | :    :    :| : |: :: |   |||:| |

Db       1199 KIIKLIRTSNPNDSLGQIIVM---DSIGNNCTMNFQNNDGS--NIGLLGFH--------- 1244

Qy        432 EEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
                  |    | ||   ::|       || | || |: || |
Db       1245 ---SDDLVASSWYYNHIRRN--TSSNGCFWSFISKEHGWKE 1280


RESULT 10
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;

  Query Match             29.0%;  Score 738;  DB 2;  Length 361;
  Best Local Similarity   94.6%;  Pred. No. 1e-30;
  Matches  139;  Conservative    7;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANS 88
              |||::||||||||||||||||:|||||||||||||||||:|||||||||||||:|
Db        215 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 274

Qy         89 KIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||:|||||||||||||||:|| ||||||||||||||||||||||||||||
Db        275 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 334

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||||||||
Db        335 GNRIIWTLIDINGKTKSVFFEYNIRED 361


RESULT 11
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
```

```
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;

  Query Match             29.0%;  Score 738;  DB 2;  Length 361;
  Best Local Similarity   94.6%;  Pred. No. 1e-30;
  Matches  139;  Conservative    7;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         29 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANS 88
              |||::|||||||||||||||||:||||||||||||||||||:|||||||||||||:|
Db        215 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLNDKNQFKLTSSADS 274

Qy         89 KIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIR 148
              ||||||||||||||:||||||||||||:|| ||||||||||||||||||||||||
Db        275 KIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIR 334

Qy        149 GNRIIWTLIDINGKTKSVFFEYNIRED 175
              |||||||||||||||||||||||||||
Db        335 GNRIIWTLIDINGKTKSVFFEYNIRED 361


RESULT 12
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;
```

```
  Query Match             27.7%;  Score 704;  DB 2;  Length 1251;
  Best Local Similarity   34.9%;  Pred. No. 2.2e-28;
  Matches  157;  Conservative   92;  Mismatches  165;  Indels   36;  Gaps   12;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  |||:   | :: :|||:|||::  :| ||| : : :   : :   :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFN 892

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
              |  |:: ::||  ||::: : :||:|||:||| | |  :   |:||||||||: |||||
Db        893 SKPSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--INNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
              |:|:  | ||||| |     : : |:|     ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGH 1010

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
              |      | ::  :  :  |:||:  :   |:| ::||:||: || :: |:  |  : :
Db       1011 LIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
              ||||||| |:|:| ||: |     |::|  :|||       | |  |  |  :    ||
Db       1070 NILKDFWGNYLLYDKGYYLLNVLKPNNFIDRRKDS------TLSINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDS 382
              | |  |:|  |  | ||  ||| | :|::: : :  : :|      :|: :      | |
Db       1124 GIKVKIQR-VNDSSTNDRFVRKNDQVYINYISNSSSYSLYADTNTTDKEKTI---KSSSS 1179

Qy        383 DEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCISK 442
                :| : :         :| : || :  : : |||:|                |    |
Db       1180 GNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAST 1222

Qy        443 WYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||   :  : :    || | || :: || |
Db       1223 WYYTHM--RDHTNSNGCFWNFISEEHGWQE 1250
```

```
RESULT 13
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
```

```
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match             27.3%;  Score 695.5;  DB 2;  Length 1252;
  Best Local Similarity   35.3%;  Pred. No. 5.9e-28;
  Matches  159;  Conservative   91;  Mismatches  164;  Indels   37;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
              | |||  |||:   | :: :||:|||::  :| ||| : : :   |     :|||| : :
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
              |:: ::||  ||::: : :||:|||:||| | |  :   ::||||||||: |||||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
              |:|:  | ||||| | :|  : : | |     ||:|||:| ||||||: | ::|:||||
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
              |      | ::  :  :  |:||:  :   |::| ::||:||: || :: |:  |  : :
Db       1011 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
              ||||||| |:|:|||||: |     |::|  : ||       | |  |  |  :    ||
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS------TLSINNIRSTILLANRLYS 1123

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTYKYFKKEEEKLFLAPISD 381
              | |  |:| :|| | ||::||| | :|::|  :      :|      :|: :    ||
Db       1124 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTI---KISS 1179

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS 441
              |   :| : :         :| : || :  : : |||:|                |  |
Db       1180 SGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1222

Qy        442 KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
              ||   ::    :   |  | || :: || |
Db       1223 TWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1251
```

```
RESULT 14
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
```

```
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match           27.3%;  Score 695.5;  DB 2;  Length 1255;
  Best Local Similarity  35.3%;  Pred. No. 6e-28;
  Matches  159;  Conservative   91;  Mismatches  164;  Indels   37;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTS 84
Db        836 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 895

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMK-NNSGW 143
Db        896 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 953

Qy        144 KISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGK 202
Db        954 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1013

Qy        203 LESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYS 262
Db       1014 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1072

Qy        263 EYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 322
Db       1073 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS------TLSINNIRSTILLANRLYS 1126

Qy        323 GEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTYKYFKKEEEKLFLAPISD 381
Db       1127 GIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTI---KISS 1182

Qy        382 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS 441
Db       1183 SGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLGFK-----------ADTVVAS 1225

Qy        442 KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 472
Db       1226 TWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1254


RESULT 15
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match           24.2%;  Score 616.5;  DB 12;  Length 1275;
  Best Local Similarity  33.8%;  Pred. No. 7.1e-24;
  Matches  159;  Conservative   78;  Mismatches  162;  Indels   71;  Gaps   13;

Qy         27 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 84
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         85 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 144
Db        910 SGD-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWK 963

Qy        145 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 203
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        204 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 263
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        264 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 323
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVQYPDRSK------LYTG 1132

Qy        324 EKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSD 383
Db       1133 NPITIKSVSDKNPYSR--ILNGDNIILHMLYNSRKYMI----------------IRDTD 1173

Qy        384 EFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV-----------F 431
Db       1174 TIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSF 1224

Qy        432 EE--------YKDYFCISKWYLKEVKRKPYNLKL---GCNWQFIPKDEGW 470
Db       1225 RENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 1274


Search completed: September 10, 2004, 03:41:28
Job time : 22.5963 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 73.2397 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 1250
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	519	41.5	804	5	ABG69083	Abg69083 Botulinum
2	368	29.4	451	2	AAW68395	Aaw68395 Clostridi
3	291	23.3	382	4	AAB04100	Aab04100 Botulism
4	228	18.2	427	5	ABG69082	Abg69082 Botulinum
5	227	18.2	422	5	ABG69071	Abg69071 Botulinum
6	79	6.3	449	3	AAY77137	Aay77137 Synthetic
7	79	6.3	449	4	AAB04094	Aab04094 Botulism
8	75	6.0	452	2	AAW68396	Aaw68396 Clostridi
9	68	5.4	419	4	AAB04095	Aab04095 Botulism
10	20	1.6	408	4	AAB04101	Aab04101 Botulism
11	20	1.6	858	5	ABG69085	Abg69085 Botulinum
12	20	1.6	1059	3	AAY93309	Aay93309 A mangane
13	20	1.6	1084	3	AAY93312	Aay93312 A mangane
14	15	1.2	144	2	AAW09016	Aaw09016 Immunogen
15	15	1.2	431	2	AAW09014	Aaw09014 Immunogen
16	15	1.2	432	3	AAY77138	Aay77138 Synthetic
17	15	1.2	432	4	AAB04103	Aab04103 Botulism
18	15	1.2	432	4	AAB04096	Aab04096 Botulism
19	15	1.2	645	4	AAE07894	Aae07894 Modified
20	15	1.2	645	6	AAE35692	Aae35692 DipT HN d
21	15	1.2	657	6	AAE35693	Aae35693 BoNT/F-Hc
22	15	1.2	657	6	AAE35694	Aae35694 BoNT/F-Hc
23	15	1.2	660	4	AAE07898	Aae07898 Modified
24	15	1.2	685	4	AAE07893	Aae07893 Modified
25	15	1.2	862	4	AAE07890	Aae07890 Modified
26	15	1.2	887	4	AAE07892	Aae07892 Modified
27	15	1.2	979	6	AAE35713	Aae35713 BoNT/F-Hc
28	15	1.2	1032	4	AAE07901	Aae07901 C. botuli
29	15	1.2	1092	4	AAE07900	Aae07900 C. botuli
30	15	1.2	1192	6	AAE35711	Aae35711 BoNT/F-Hc
31	15	1.2	1192	6	AAE35710	Aae35710 BoNT/F-Hc
32	12	1.0	439	5	ABG69073	Abg69073 Botulinum
33	12	1.0	444	5	ABG69086	Abg69086 Botulinum
34	12	1.0	448	2	AAW68399	Aaw68399 Clostridi
35	12	1.0	848	5	ABG69087	Abg69087 Botulinum
36	11	0.9	144	2	AAW09015	Aaw09015 Immunogen
37	11	0.9	436	5	ABG69072	Abg69072 Botulinum
38	11	0.9	443	5	ABG69084	Abg69084 Botulinum
39	10	0.8	233	3	AAY77143	Aay77143 Botulinum
40	10	0.8	382	3	AAB36303	Aab36303 BoNT/A pr
41	10	0.8	415	4	AAB04083	Aab04083 Botulism
42	10	0.8	432	3	AAY77142	Aay77142 Native bo
43	10	0.8	434	4	AAB04089	Aab04089 Botulism
44	10	0.8	435	4	AAB04090	Aab04090 Botulism
45	10	0.8	437	4	AAB04088	Aab04088 Botulism

ALIGNMENTS

RESULT 1
ABG69083
ID ABG69083 standard; protein; 804 AA.
XX
AC ABG69083;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #17.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98553.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 152-154; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

```
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 804 AA;

  Query Match             41.5%;  Score 519;  DB 5;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 2
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             29.4%;  Score 368;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  368;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        827 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 86

Qy        887 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 146

Qy        947 NSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 NSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 206

Qy       1007 INGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNE 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 INGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNE 266

Qy       1067 PNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIK 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 PNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIK 326

Qy       1127 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQ 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQ 386

Qy       1187 VVVMNSVG 1194
              ||||||||
Db        387 VVVMNSVG 394


RESULT 3
AAB04100
ID   AAB04100 standard; protein; 382 AA.
XX
AC   AAB04100;
```

XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain N-terminal sequence (serotype E).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54496.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 6; Fig 15b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 382 AA;

Query Match 23.3%; Score 291; DB 4; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.4e-300;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        424 SICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGL 483
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGL 61

Qy        484 SDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDT 543
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 SDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDT 121

Qy        544 ALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVV 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVV 181
```

```
Qy        604 PYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKV 663
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 PYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKV 241

Qy        664 IKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 IKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 292
```

RESULT 4
ABG69082
ID ABG69082 standard; protein; 427 AA.
XX
AC ABG69082;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #16.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98552.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 150-151; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 427 AA;

Query Match 18.2%; Score 228; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.3e-233;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRK 423
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRK 423
```

RESULT 5
ABG69071
ID ABG69071 standard; protein; 422 AA.
XX
AC ABG69071;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #5.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98541.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 125-126; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 422 AA;

```
  Query Match           18.2%;  Score 227;  DB 5;  Length 422;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-232;
  Matches  227;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIR 422
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIR 422
```

RESULT 6
AAY77137
ID AAY77137 standard; protein; 449 AA.
XX
AC AAY77137;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype E (BoNTE) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype E; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87215.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 26; Page 43-44; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and

```
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 449 AA;

  Query Match             6.3%;  Score 79;  DB 3;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-74;
  Matches   79;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        817 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 876
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 73

Qy        877 DVYKYPTNKNQFGIYNDKL 895
              |||||||||||||||||||
Db         74 DVYKYPTNKNQFGIYNDKL 92


RESULT 7
AAB04094
ID   AAB04094 standard; protein; 449 AA.
XX
AC   AAB04094;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54488.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 7b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 449 AA;

  Query Match             6.3%;  Score 79;  DB 4;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-74;
  Matches   79;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        817 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 876
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 73

Qy        877 DVYKYPTNKNQFGIYNDKL 895
              |||||||||||||||||||
Db         74 DVYKYPTNKNQFGIYNDKL 92


RESULT 8
AAW68396
ID   AAW68396 standard; protein; 452 AA.
XX
AC   AAW68396;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain NCTC 11219.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
```

CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 452 AA;

Query Match 6.0%; Score 75; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 2.9e-70;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 892 NDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWK 951
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 92 NDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWK 151

Qy 952 VSLNHNEIIWTLQDN 966
|||||||||||||||
Db 152 VSLNHNEIIWTLQDN 166

RESULT 9
AAB04095
ID AAB04095 standard; protein; 419 AA.
XX
AC AAB04095;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54489.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Fig 8; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 419 AA;

Query Match 5.4%; Score 68; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 7.7e-63;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 899 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 958
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 125

Qy 959 IIWTLQDN 966
||||||||
Db 126 IIWTLQDN 133

RESULT 10
AAB04101
ID AAB04101 standard; protein; 408 AA.
XX
AC AAB04101;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain N-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54497.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 6; Fig 16b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino

CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 408 AA;

Query Match 1.6%; Score 20; DB 4; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 273 INTQFNKRKEQMYQALQNQV 292

RESULT 11
ABG69085
ID ABG69085 standard; protein; 858 AA.
XX
AC ABG69085;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #19.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98555.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 157-159; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 858 AA;

Query Match 1.6%; Score 20; DB 5; Length 858;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 723 INTQFNKRKEQMYQALQNQV 742

RESULT 12
AAY93309
ID AAY93309 standard; protein; 1059 AA.
XX
AC AAY93309;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 48-51; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype F. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a domain
CC that translocates the SOD of the composition into the neuronal cell.
CC After translocation, the linker is cleaved to release the SOD. The
CC composition is useful for treating neuronal diseases caused or augmented
CC by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC disease, Huntington's disease and motor neurone diseases
XX
SQ Sequence 1059 AA;

Query Match 1.6%; Score 20; DB 3; Length 1059;

```
  Best Local Similarity   100.0%;  Pred. No. 2.7e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        493 INTQFNKRKEQMYQALQNQV 512


RESULT 13
AAY93312
ID   AAY93312 standard; protein; 1084 AA.
XX
AC   AAY93312;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 57-60; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide dismutase
CC   (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC   cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC   serotype F. The specification describes a composition for delivery of SOD
CC   to neuronal cells. The composition comprises SOD linked, by a cleavable
CC   linker, to a neuronal cell targeting component (NCTC). This component has
CC   a domain that binds to a neuronal cell and a domain that translocates the
CC   SOD of the composition into the neuronal cell. After translocation, the
CC   linker is cleaved to release the SOD. The composition is useful for
CC   treating neuronal diseases caused or augmented by oxidative stress, such
CC   as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases
XX
SQ   Sequence 1084 AA;

  Query Match             1.6%;  Score 20;  DB 3;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        518 INTQFNKRKEQMYQALQNQV 537


RESULT 14
AAW09016
ID   AAW09016 standard; protein; 144 AA.
XX
AC   AAW09016;
XX
DT   17-OCT-2003  (revised)
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin polypeptide (aa992-1135).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum; type F strain Langeland.
XX
PN   WO9641881-A1.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-GB001409.
XX
PR   12-JUN-1995;   95GB-00011909.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
XX
DR   WPI; 1997-065467/06.
XX
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT   vaccine prodn.
XX
PS   Claim 5; Page 18-19; 37pp; English.
XX
CC   Novel polypeptides (AAW09014-17) respectively comprise amino acids 848-
CC   1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a type F
CC   botulinum neurotoxin (BoNT/F). They lack the L chain and HN epitopes
CC   necessary for metalloprotease activity and toxin internalisation. They
CC   are free of botulinum toxin activity but can induce protective immunity
CC   to a type F botulinum toxin, making them useful for vaccine prodn.
CC   Recombinant polypeptides can be produced in transformed host cells, esp.
CC   as fusion proteins, e.g. with maltose binding protein to facilitate
CC   purification. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 144 AA;

  Query Match             1.2%;  Score 15;  DB 2;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        983 ISDYINKWIFVTITN 997
              |||||||||||||||
Db         10 ISDYINKWIFVTITN 24


RESULT 15
AAW09014
ID   AAW09014 standard; protein; 431 AA.
XX
AC   AAW09014;
XX
DT   17-OCT-2003  (revised)
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum; type F strain Langeland.
XX
PN   WO9641881-A1.
XX
```

PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB001409.
XX
PR 12-JUN-1995; 95GB-00011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX
CC A polypeptide (AAW09014) comprises the heavy chain (amino acids 848-1278)
CC of a type F botulinum neurotoxin (BoNT/F), and can be produced using a
CC synthetic gene (AAT48101) based on the natural gene sequence (AAT48100)
CC for the heavy chain. The polypeptides and its fragments (see also
CC AAW09015-17) lack the light chain and HN epitopes necessary for
CC metalloprotease activity and toxin internalisation. They are free of
CC botulinum toxin activity but can induce protective immunity to a type F
CC botulinum toxin, making them useful for vaccine prodn. Recombinant
CC polypeptides can be produced in transformed host cells, esp. as fusion
CC proteins, e.g. with maltose binding protein to facilitate purification.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 431 AA;

Query Match 1.2%; Score 15; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        983 ISDYINKWIFVTITN 997
              |||||||||||||||
Db        154 ISDYINKWIFVTITN 168
```

Search completed: September 10, 2004, 05:05:56
Job time : 74.2397 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 20.4944 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 1250
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	0.8	438	1	US-08-480-604A-23	Sequence 23, Appl
2	10	0.8	438	2	US-08-405-496A-23	Sequence 23, Appl
3	10	0.8	438	3	US-08-915-136-23	Sequence 23, Appl
4	10	0.8	438	4	US-09-084-517-23	Sequence 23, Appl
5	10	0.8	462	1	US-08-480-604A-26	Sequence 26, Appl
6	10	0.8	462	2	US-08-405-496A-26	Sequence 26, Appl
7	10	0.8	462	3	US-08-915-136-26	Sequence 26, Appl
8	10	0.8	462	4	US-09-084-517-26	Sequence 26, Appl
9	10	0.8	1296	1	US-08-480-604A-28	Sequence 28, Appl
10	10	0.8	1296	2	US-08-405-496A-28	Sequence 28, Appl
11	10	0.8	1296	3	US-08-915-136-28	Sequence 28, Appl
12	10	0.8	1296	4	US-09-084-517-28	Sequence 28, Appl
13	9	0.7	858	4	US-09-255-829-22	Sequence 22, Appl
14	9	0.7	858	4	US-09-255-829-29	Sequence 29, Appl
15	9	0.7	1169	4	US-09-255-829-20	Sequence 20, Appl
16	8	0.6	466	4	US-09-134-001C-3355	Sequence 3355, Ap
17	8	0.6	473	4	US-09-252-991A-28845	Sequence 28845, A
18	8	0.6	793	4	US-09-107-532A-6223	Sequence 6223, Ap
19	7	0.6	14	2	US-08-726-306A-77	Sequence 77, Appl
20	7	0.6	17	1	US-08-149-809-2	Sequence 2, Appli
21	7	0.6	35	4	US-09-266-764-7	Sequence 7, Appli
22	7	0.6	68	4	US-09-134-000C-4208	Sequence 4208, Ap
23	7	0.6	79	2	US-09-196-048-2	Sequence 2, Appli
24	7	0.6	177	4	US-09-252-991A-29795	Sequence 29795, A
25	7	0.6	199	4	US-09-134-001C-3825	Sequence 3825, Ap
26	7	0.6	199	4	US-09-134-001C-4609	Sequence 4609, Ap
27	7	0.6	283	4	US-09-198-452A-803	Sequence 803, App
28	7	0.6	329	4	US-09-740-035-2	Sequence 2, Appli
29	7	0.6	335	4	US-09-489-039A-11298	Sequence 11298, A
30	7	0.6	339	4	US-09-252-991A-21904	Sequence 21904, A
31	7	0.6	359	4	US-09-403-343B-21	Sequence 21, Appl
32	7	0.6	379	4	US-09-740-035-4	Sequence 4, Appli
33	7	0.6	411	4	US-09-328-352-6719	Sequence 6719, Ap
34	7	0.6	417	4	US-09-252-991A-16876	Sequence 16876, A
35	7	0.6	434	4	US-09-252-991A-23499	Sequence 23499, A
36	7	0.6	438	4	US-09-134-000C-4760	Sequence 4760, Ap
37	7	0.6	452	1	US-07-618-312A-2	Sequence 2, Appli
38	7	0.6	452	1	US-07-618-312A-4	Sequence 4, Appli
39	7	0.6	452	1	US-08-110-786A-8	Sequence 8, Appli
40	7	0.6	452	1	US-08-280-228-2	Sequence 2, Appli
41	7	0.6	452	1	US-08-280-228-4	Sequence 4, Appli
42	7	0.6	472	4	US-09-328-352-6296	Sequence 6296, Ap
43	7	0.6	502	4	US-09-796-202-16	Sequence 16, Appl
44	7	0.6	529	1	US-08-484-105-8	Sequence 8, Appli
45	7	0.6	529	1	US-08-484-106-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match             0.8%;  Score 10;  DB 1;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 2
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match             0.8%;  Score 10;  DB 2;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 3
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match             0.8%;  Score 10;  DB 3;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
```

```
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 4
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-23

  Query Match             0.8%;  Score 10;  DB 4;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 5
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             0.8%;  Score 10;  DB 1;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0.61;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        125 NNEYTIINCM 134


RESULT 6
US-08-405-496A-26
; Sequence 26, Application US/08405496A
```

```
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match             0.8%;  Score 10;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0.61;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        125 NNEYTIINCM 134


RESULT 7
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             0.8%;  Score 10;  DB 3;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0.61;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        125 NNEYTIINCM 134


RESULT 8
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             0.8%;  Score 10;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0.61;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        125 NNEYTIINCM 134


RESULT 9
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             0.8%;  Score 10;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 10
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
```

```
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             0.8%;  Score 10;  DB 2;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 11
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             0.8%;  Score 10;  DB 3;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 12
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             0.8%;  Score 10;  DB 4;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 13
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-22

  Query Match             0.7%;  Score 9;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        621 VPYIGLALN 629


RESULT 14
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             0.7%;  Score 9;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        621 VPYIGLALN 629


RESULT 15
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
```

; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 0.7%; Score 9; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 VPYIGLALN 611
|||||||||
Db 621 VPYIGLALN 629

Search completed: September 10, 2004, 05:17:06
Job time : 21.4944 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 63.2308 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 1250
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1250	100.0	1250	12	US-10-272-898-50	Sequence 50, Appl
2	1250	100.0	1250	15	US-10-354-774-50	Sequence 50, Appl
3	1250	100.0	1250	15	US-10-271-012-50	Sequence 50, Appl
4	1250	100.0	1250	16	US-10-729-122-50	Sequence 50, Appl
5	1250	100.0	1250	16	US-10-729-039-50	Sequence 50, Appl
6	1093	87.4	1251	15	US-10-452-024-5	Sequence 5, Appli
7	1093	87.4	1251	15	US-10-452-024-124	Sequence 124, App
8	1093	87.4	1252	15	US-10-452-024-125	Sequence 125, App
9	1093	87.4	1252	15	US-10-205-516-10	Sequence 10, Appl
10	1093	87.4	1255	15	US-10-452-024-126	Sequence 126, App
11	793	63.4	1262	15	US-10-205-516-24	Sequence 24, Appl
12	519	41.5	804	13	US-10-011-588-37	Sequence 37, Appl
13	519	41.5	1252	12	US-10-272-898-52	Sequence 52, Appl
14	519	41.5	1252	15	US-10-354-774-52	Sequence 52, Appl
15	519	41.5	1252	15	US-10-271-012-52	Sequence 52, Appl
16	519	41.5	1252	15	US-10-452-024-128	Sequence 128, App
17	519	41.5	1252	16	US-10-729-122-52	Sequence 52, Appl
18	519	41.5	1252	16	US-10-729-039-52	Sequence 52, Appl
19	441	35.3	1251	15	US-10-452-024-127	Sequence 127, App
20	415	33.2	1251	15	US-10-452-024-129	Sequence 129, App
21	368	29.4	451	12	US-10-272-898-54	Sequence 54, Appl
22	368	29.4	451	15	US-10-354-774-54	Sequence 54, Appl
23	368	29.4	451	15	US-10-271-012-54	Sequence 54, Appl
24	368	29.4	451	16	US-10-729-122-54	Sequence 54, Appl
25	368	29.4	451	16	US-10-729-039-54	Sequence 54, Appl
26	291	23.3	382	10	US-09-910-186A-28	Sequence 28, Appl
27	236	18.9	438	15	US-10-452-024-143	Sequence 143, App
28	229	18.3	252	15	US-10-452-024-148	Sequence 148, App
29	228	18.2	427	13	US-10-011-588-35	Sequence 35, Appl
30	227	18.2	422	13	US-10-011-588-13	Sequence 13, Appl
31	150	12.0	449	10	US-09-910-186A-14	Sequence 14, Appl
32	140	11.2	452	12	US-10-272-898-56	Sequence 56, Appl
33	140	11.2	452	15	US-10-354-774-56	Sequence 56, Appl
34	140	11.2	452	15	US-10-271-012-56	Sequence 56, Appl
35	140	11.2	452	16	US-10-729-122-56	Sequence 56, Appl
36	140	11.2	452	16	US-10-729-039-56	Sequence 56, Appl
37	107	8.6	438	15	US-10-452-024-147	Sequence 147, App
38	52	4.2	252	15	US-10-452-024-155	Sequence 155, App
39	20	1.6	366	15	US-10-452-024-173	Sequence 173, App
40	20	1.6	369	15	US-10-452-024-174	Sequence 174, App
41	20	1.6	408	10	US-09-910-186A-30	Sequence 30, Appl
42	20	1.6	858	13	US-10-011-588-41	Sequence 41, Appl
43	20	1.6	1268	15	US-10-452-024-156	Sequence 156, App
44	20	1.6	1274	12	US-10-272-898-71	Sequence 71, Appl
45	20	1.6	1274	15	US-10-354-774-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-50
; Sequence 50, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-272-898-50

  Query Match             100.0%;  Score 1250;  DB 12;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 2
US-10-354-774-50
; Sequence 50, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-354-774-50

  Query Match             100.0%;  Score 1250;  DB 15;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
```

```
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 3
US-10-271-012-50
; Sequence 50, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-271-012-50

  Query Match           100.0%;  Score 1250;  DB 15;  Length 1250;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
```

```
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

RESULT 4
```

```
US-10-729-122-50
; Sequence 50, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-729-122-50

  Query Match           100.0%;  Score 1250;  DB 16;  Length 1250;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 5
US-10-729-039-50
; Sequence 50, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-729-039-50

  Query Match             100.0%;  Score 1250;  DB 16;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
```

```
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
```

RESULT 6
US-10-452-024-5
; Sequence 5, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-5

Query Match 87.4%; Score 1093; DB 15; Length 1251;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy           1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Db           1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy          61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db          61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy         121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db         121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy         181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db         181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy         241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db         241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy         301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db         301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy         361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db         361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy         421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db         421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy         481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db         481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy         541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db         541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy         601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db         601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy         661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db         661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy         721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db         721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy         781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db         781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy         841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db         841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy         901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db         901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy         961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db         961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy        1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db        1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
```

```
Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194


RESULT 7
US-10-452-024-124
; Sequence 124, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 124
;   LENGTH: 1251
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-124

  Query Match            87.4%;  Score 1093;  DB 15;  Length 1251;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194


RESULT 8
US-10-452-024-125
; Sequence 125, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 125
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-125

  Query Match            87.4%;  Score 1093;  DB 15;  Length 1252;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
```

```
Qy        61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
Db        61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy       121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db       121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy       181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Db       181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy       241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db       241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy       301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db       301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy       361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db       361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy       421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db       421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy       481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db       481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy       541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db       541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy       601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db       601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy       661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
Db       661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy       721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db       721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy       781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
Db       781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy       841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db       841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy       901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db       901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy       961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
Db       961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy      1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
Db      1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy      1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Db      1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy      1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
Db      1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
```

RESULT 9
US-10-205-516-10
; Sequence 10, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1252
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-10

Query Match 87.4%; Score 1093; DB 15; Length 1252;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
Db         1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy        61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
Db        61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy       121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db       121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy       181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Db       181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy       241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db       241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy       301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db       301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy       361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db       361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy       421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db       421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy       481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db       481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy       541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db       541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy       601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db       601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy       661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
```

```
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194


RESULT 10
US-10-452-024-126
; Sequence 126, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 126
;   LENGTH: 1255
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-126

  Query Match           87.4%;  Score 1093;  DB 15;  Length 1255;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 63

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 123

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 183

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 243

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 303

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 363

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 423

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 483

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 543

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 603

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 663

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 723

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 783

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 843

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 903

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 963

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1023

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1083

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1143

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1197


RESULT 11
US-10-205-516-24
; Sequence 24, Application US/10205516
; Publication No. US20040018589A1
```

; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 1262
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-24

Query Match 63.4%; Score 793; DB 15; Length 1262;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1193; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

```
Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 I----RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFN 476
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRIEGRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFN 480

Qy        477 SESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVN 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVN 540

Qy        537 LTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKI 596
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        541 LTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKI 600

Qy        597 ADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGS 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGS 660

Qy        657 SDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALK 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALK 720

Qy        717 AIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINE 780

Qy        777 VKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKIL 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 VKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKIL 840

Qy        837 ISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS 900

Qy        897 EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNH 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNH 960

Qy        957 NEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKS 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 NEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKS 1020

Qy       1017 ILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFW 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFW 1080

Qy       1077 GNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS 1140

Qy       1137 TNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1198
```

RESULT 12
US-10-011-588-37
; Sequence 37, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 804
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:36
US-10-011-588-37

Query Match 41.5%; Score 519; DB 13; Length 804;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
```

```
Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 13
US-10-272-898-52
; Sequence 52, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1252 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-272-898-52

  Query Match             41.5%;  Score 519;  DB 12;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 14
US-10-354-774-52
; Sequence 52, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
```

```
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 1252 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-354-774-52

  Query Match             41.5%;  Score 519;  DB 15;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 15
US-10-271-012-52
; Sequence 52, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;        COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/271,012
;            FILING DATE: 15-Oct-2002
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 1252 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-271-012-52

  Query Match             41.5%;  Score 519;  DB 15;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


Search completed: September 10, 2004, 05:52:39
Job time : 67.2308 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 17.317 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 1250
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1093	87.4	1251	2	JH0256	botulinum neurotox
2	519	41.5	1252	2	S21178	botulinum neurotox
3	150	12.0	367	2	S48106	neurotoxin type E
4	44	3.5	122	2	A53878	type E neurotoxin
5	20	1.6	366	2	S48110	neurotoxin type F
6	20	1.6	369	2	S48109	neurotoxin type F
7	20	1.6	1268	2	S33411	botulinum neurotox
8	20	1.6	1274	2	I40813	neurotoxin type F
9	12	1.0	1297	2	S39791	neurotoxin - Clost
10	10	0.8	1296	1	BTCLAB	bontoxilysin (EC 3
11	10	0.8	1296	2	I40645	botulinum neurotox
12	9	0.7	1291	1	A48940	bontoxilysin (EC 3
13	9	0.7	1291	2	I40631	non-proteolytic bo
14	8	0.6	13	2	S08575	botulinum neurotox
15	8	0.6	162	2	G70005	hypothetical prote
16	8	0.6	397	2	E97176	N-terminal domain
17	8	0.6	425	2	T38536	probable DNA-direc
18	8	0.6	437	2	C95393	probable transport
19	8	0.6	440	2	T51556	sugar transporter-
20	8	0.6	456	2	A83264	hypothetical prote
21	8	0.6	535	2	T37508	hypothetical prote
22	8	0.6	769	2	S50966	probable membrane
23	8	0.6	773	2	C84554	hypothetical prote
24	7	0.6	65	2	AF3177	conserved hypothet
25	7	0.6	68	2	H85826	hypothetical prote
26	7	0.6	73	2	H71149	hypothetical prote
27	7	0.6	73	2	S10650	hypothetical prote
28	7	0.6	82	2	T07144	probable RNA-direc
29	7	0.6	95	2	C86685	prophage pi1 prote
30	7	0.6	95	2	B86757	prophage pi2 prote
31	7	0.6	95	2	T13259	holin protein homo
32	7	0.6	99	2	D86797	prophage pi3 prote
33	7	0.6	112	2	D95269	hypothetical prote
34	7	0.6	117	2	H90317	conserved hypothet
35	7	0.6	118	2	C97927	hypothetical prote
36	7	0.6	118	2	B95058	transcription regu
37	7	0.6	119	2	S53460	hypothetical prote
38	7	0.6	120	1	TTMSA	glycoprotein hormo
39	7	0.6	123	2	A39266	heat shock protein
40	7	0.6	123	2	B71665	hypothetical prote
41	7	0.6	123	2	C97814	hypothetical prote
42	7	0.6	123	2	E84752	hypothetical prote
43	7	0.6	141	2	H64909	ydfO protein - Esc
44	7	0.6	145	2	S27620	replication protei
45	7	0.6	151	2	F83689	hypothetical prote

ALIGNMENTS

RESULT 1
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.,
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxir
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release a
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 87.4%; Score 1093; DB 2; Length 1251;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194


RESULT 2
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin,
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match             41.5%;  Score 519;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
```

```
Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
```

RESULT 3
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

```
  Query Match            12.0%;  Score 150;  DB 2;  Length 367;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-147;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        817 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 261

Qy        877 DVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 DVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE 321

Qy        937 YTIINCMRDNNSGWKVSLNHNEIIWTLQDN 966
              ||||||||||||||||||||||||||||||
Db        322 YTIINCMRDNNSGWKVSLNHNEIIWTLQDN 351
```

RESULT 4
A53878
type E neurotoxin - Clostridium botulinum (fragments)
C;Species: Clostridium botulinum
C;Date: 27-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C;Accession: A53878
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 11, 255-264, 1992
A;Title: Pepsin fragmentation of botulinum type E neurotoxin: isolation and characteriza
A;Reference number: A53878; MUID:93000392; PMID:1388670
A;Contents: type E, E-43, Alaska
A;Accession: A53878
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-122 <GIM>
A;Note: sequence modified after extraction from NCBI backbone
A;Note: sequence extracted from NCBI backbone (NCBIP:115403, NCBIP:115404)
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

```
  Query Match             3.5%;  Score 44;  DB 2;  Length 122;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-37;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1116 LLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHL 1159
              |||||||||||||||||||||||||||||||||||||||||||||
Db         62 LLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHL 105
```

RESULT 5
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

```
  Query Match             1.6%;  Score 20;  DB 2;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db         78 INTQFNKRKEQMYQALQNQV 97
```

RESULT 6
S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin

```
  Query Match             1.6%;  Score 20;  DB 2;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db         78 INTQFNKRKEQMYQALQNQV 97
```

RESULT 7
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999

C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 1.6%; Score 20; DB 2; Length 1268;
Best Local Similarity 100.0%; Pred. No. 1.6e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 702 INTQFNKRKEQMYQALQNQV 721

RESULT 8
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 1.6%; Score 20; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 1.6e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 711 INTQFNKRKEQMYQALQNQV 730

RESULT 9
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 1.0%; Score 12; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 DPALTLMHELIH 216
||||||||||||
Db 223 DPALTLMHELIH 234

RESULT 10
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.;
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin,
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-terma
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ext
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arour
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987

A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 0.8%; Score 10; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 934 NNEYTIINCM 943
||||||||||
Db 959 NNEYTIINCM 968

RESULT 11
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 0.8%; Score 10; DB 2; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 946 NNSGWKVSLN 955
||||||||||
Db 970 NNSGWKVSLN 979

RESULT 12
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisio
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

```
  Query Match             0.7%;  Score 9;  DB 1;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        621 VPYIGLALN 629


RESULT 13
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match             0.7%;  Score 9;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        621 VPYIGLALN 629


RESULT 14
S08575
botulinum neurotoxin type E - Clostridium botulinum (strain Alaska E-43) (fragment)
C;Species: Clostridium botulinum
A;Variety: strain Alaska E-43
C;Date: 19-Mar-1997 #sequence_revision 21-Nov-1998 #text_change 21-Nov-1998
C;Accession: S08575
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S08575
A;Molecule type: protein
A;Residues: 1-13 <SCH>
A;Experimental source: strain Alaska E-43
F;1-13/Product: botulinum neurotoxin E light chain (fragment) #status predicted <LIG>

  Query Match             0.6%;  Score 8;  DB 2;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 0.88;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 INSFNYND 11
              ||||||||
Db          3 INSFNYND 10


RESULT 15
G70005
hypothetical protein yuaE - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: G70005
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berterc
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G70005
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-162 <KUN>
A;Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB15081.1; PID:e1185976;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yuaE

  Query Match             0.6%;  Score 8;  DB 2;  Length 162;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        482 GLSDEKLN 489
              ||||||||
Db         17 GLSDEKLN 24


Search completed: September 10, 2004, 05:14:48
Job time : 20.417 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 10.8033 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 1250
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1092	87.4	1250	1	BXE_CLOBU	P30995 clostridium
2	415	33.2	1250	1	BXE_CLOBO	Q00496 clostridium
3	20	1.6	1274	1	BXF_CLOBO	P30996 clostridium
4	12	1.0	1296	1	BXG_CLOBO	Q60393 clostridium
5	10	0.8	1295	1	BXA1_CLOBO	P10845 clostridium
6	10	0.8	1295	1	BXA2_CLOBO	Q45894 clostridium
7	9	0.7	550	1	RPOA_CHLRE	Q8huh2 chlamydomon
8	9	0.7	1290	1	BXB_CLOBO	P10844 clostridium
9	8	0.6	425	1	RPA3_SCHPO	O14086 schizosacch
10	8	0.6	838	1	HIS2_CANAL	O74712 candida alb
11	8	0.6	980	1	SYN_DROME	Q24546 drosophila
12	7	0.6	73	1	TRM1_PYRWO	P20300 pyrococcus
13	7	0.6	120	1	GLH2_RAT	P11963 r glycoprot
14	7	0.6	120	1	GLHA_MASCO	Q9erg4 m glycoprot
15	7	0.6	120	1	GLHA_MERUN	Q9erj6 m glycoprot
16	7	0.6	120	1	GLHA_MOUSE	P01216 m glycoprot
17	7	0.6	123	1	Y600_RICPR	Q9zcv8 rickettsia
18	7	0.6	123	1	Y915_RICCN	Q9akn7 rickettsia
19	7	0.6	136	1	YDFO_ECOLI	P76156 escherichia
20	7	0.6	160	1	RL11_METTH	O27715 methanobact
21	7	0.6	161	1	LI52_CAEEL	Q10120 caenorhabdi
22	7	0.6	180	1	PUR6_ARCFU	O28997 archaeoglob
23	7	0.6	189	1	E2BE_CANAL	P87163 candida alb
24	7	0.6	192	1	Y3_SOCMV	P15633 soybean chl
25	7	0.6	193	1	Y455_METJA	Q57897 methanococc
26	7	0.6	204	1	Y860_RICPR	P41078 rickettsia
27	7	0.6	211	1	SN23_HUMAN	O00161 homo sapien
28	7	0.6	218	1	KAD_CLOPE	Q8xhu4 clostridium
29	7	0.6	218	1	RR3_LOTJA	Q9bbp8 lotus japon
30	7	0.6	230	1	AG11_ARATH	Q38836 arabidopsis
31	7	0.6	240	1	MYF6_XENLA	Q92020 xenopus lae
32	7	0.6	242	1	Y956_HELPJ	Q9zkp5 helicobacte
33	7	0.6	242	1	Y956_HELPY	O25610 helicobacte
34	7	0.6	243	1	FRHG_METVO	Q00393 methanococc
35	7	0.6	244	1	ARC3_CBCP	Q00901 clostridium
36	7	0.6	246	1	TRMD_LACPL	Q88wj2 lactobacill
37	7	0.6	251	1	RECO_STRCO	Q9l2h3 streptomyce
38	7	0.6	254	1	Y007_MYCGE	P47253 mycoplasma
39	7	0.6	256	1	TPIS_WIGBR	Q8d2t0 wiggleswort
40	7	0.6	274	1	TPM1_METEN	Q25456 metapenaeus
41	7	0.6	274	1	TPM1_PANST	O61379 panulirus s
42	7	0.6	304	1	LST_HAEIN	Q48211 haemophilus
43	7	0.6	305	1	HEM6_PSEAE	P43898 pseudomonas
44	7	0.6	315	1	RLUC_BUCBP	Q89ah2 buchnera ap
45	7	0.6	325	1	GPR_CLOPE	Q8xis3 clostridium

ALIGNMENTS

RESULT 1
BXE_CLOBU
ID BXE_CLOBU STANDARD; PRT; 1250 AA.
AC P30995;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43181, and ATCC 43755;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinal neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA Yokosawa N., Yashiki T., Oguma K.;
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.
RC STRAIN=5262;
RA Gimenez J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             87.4%;  Score 1092;  DB 1;  Length 1250;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1192;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 PTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 60

Qy         62 KNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTPD 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTPD 120

Qy        122 GDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 180

Qy        182 AIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 240

Qy        242 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 300

Qy        302 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 360

Qy        362 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 420

Qy        422 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 480

Qy        482 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 540

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISI 600

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 660

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 720

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 780

Qy        782 LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 840

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 900

Qy        902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 960

Qy        962 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1020

Qy       1022 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLL 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLL 1080

Qy       1082 YDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 YDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1140

Qy       1142 VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1193
```

```
RESULT 2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
```

```
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             33.2%;  Score 415;  DB 1;  Length 1250;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  515;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        199 MNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGT 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 MNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGT 257

Qy        259 DLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYS 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 DLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYS 317

Qy        319 VNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN 378
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        318 VNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN 377

Qy        379 NLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVA 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 NLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVA 437

Qy        439 SENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYI 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 SENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYI 497

Qy        499 PKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSE 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 PKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSE 557

Qy        559 FINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQK 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 FINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQK 617

Qy        619 GNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 GNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW 677

Qy        679 KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||
Db        678 KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 713


RESULT 3
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
```

```
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             1.6%;  Score 20;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        711 INTQFNKRKEQMYQALQNQV 730


RESULT 4
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             1.0%;  Score 12;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.003;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        205 DPALTLMHELIH 216
              ||||||||||||
Db        222 DPALTLMHELIH 233


RESULT 5
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
```

```
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match              0.8%;  Score 10;  DB 1;  Length 1295;
  Best Local Similarity  100.0%;  Pred. No. 0.32;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        958 NNEYTIINCM 967


RESULT 6
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
```

```
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             0.8%;  Score 10;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        946 NNSGWKVSLN 955
              ||||||||||
Db        969 NNSGWKVSLN 978


RESULT 7
RPOA_CHLRE
ID   RPOA_CHLRE     STANDARD;      PRT;   550 AA.
AC   Q8HUH2;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (PEP) (Plastid-
DE   encoded RNA polymerase alpha subunit) (RNA polymerase alpha subunit).
GN   RPOA.
OS   Chlamydomonas reinhardtii.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A., IDENTIFICATION, COMPLETE PLASTID GENOME, AND
RP   EXPRESSION OF MRNA.
RX   MEDLINE=22305394; PubMed=12417694;
RA   Maul J.E., Lilly J.W., Cui L., dePamphilis C.W., Miller W.,
RA   Harris E.H., Stern D.B.;
RT   "The Chlamydomonas reinhardtii plastid chromosome: islands of genes in
RT   a sea of repeats.";
RL   Plant Cell 14:2659-2679(2002).
CC   -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC       of DNA into RNA using the four ribonucleoside triphosphates as
CC       substrates.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: In chloroplasts the minimal PEP RNA polymerase is
CC       composed of four subunits: alpha, beta, beta', and beta'' (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- DOMAIN: The N-terminal domain is essential for RNAP assembly and
CC       basal transcription, whereas the C-terminal domain is involved in
CC       interaction with transcriptional regulators and with upstream
CC       promoter elements (By similarity).
CC   -!- SIMILARITY: Belongs to the RNA polymerase alpha chain family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF541867; AAN17816.1; -.
DR   EMBL; BK000554; DAA00934.1; -.
DR   HAMAP; MF_00059; atypical; 1.
DR   InterPro; IPR009025; RBP11-like_RNApo.
DR   InterPro; IPR001700; RNA_polA_bac_org.
DR   Pfam; PF01000; RNA_pol_A_bac; 1.
DR   ProDom; PD001179; RNA_polA_bac_org; 1.
DR   SMART; SM00662; RPOLD; 1.
KW   Transcription; Transferase; DNA-directed RNA polymerase; Chloroplast.
FT   DOMAIN        1    333       ALPHA N-TERMINAL DOMAIN (ALPHA-NTD) (BY
FT                                SIMILARITY).
FT   DOMAIN      185    258       INSERT.
FT   DOMAIN      378    550       ALPHA C-TERMINAL DOMAIN (ALPHA-CTD) (BY
FT                                SIMILARITY).
SQ   SEQUENCE   550 AA;  62290 MW;  931F910FF6359948 CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        650 IKSFLGSSD 658
              |||||||||
Db        447 IKSFLGSSD 455


RESULT 8
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
```

```
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match              0.7%;  Score 9;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 3.3;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        620 VPYIGLALN 628


RESULT 9
RPA3_SCHPO
ID   RPA3_SCHPO     STANDARD;      PRT;   425 AA.
AC   O14086;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   DNA-directed RNA polymerase I 49 kDa polypeptide (EC 2.7.7.6) (A49).
GN   RPA49 OR SPAC2F3.03C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Shematorova E.K., Shpakovski G.V.;
RT   "Molecular cloning and complementation studies of the fission yeast
RT   genes encoding subunits Rpa12 and Rpa49 specific for nuclear RNA
RT   polymerase I.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
```

```
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC       of DNA into RNA using the four ribonucleoside triphosphates as
CC       substrates. RNA polymerase I (or A) is essentially used to
CC       transcribe ribosomal DNA units.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: RNA polymerase I consists of 14 different subunits (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are
CC       found in eukaryotic nuclei: polymerase I for the ribosomal RNA
CC       precursor, polymerase II for the mRNA precursor, and polymerase
CC       III for 5S and tRNA genes.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ278613; CAC20964.1; -.
DR   EMBL; Z99165; CAB16261.1; -.
DR   PIR; T38536; T38536.
DR   GeneDB_SPombe; SPAC2F3.03c; -.
KW   Transferase; DNA-directed RNA polymerase; Transcription;
KW   Nuclear protein.
SQ   SEQUENCE   425 AA;  48066 MW;  1191EC48E6A75A76 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         78 QEKDKFLK 85
              ||||||||
Db        121 QEKDKFLK 128


RESULT 10
HIS2_CANAL
ID   HIS2_CANAL     STANDARD;      PRT;   838 AA.
AC   O74712;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Histidine biosynthesis trifunctional protein [Includes:
DE   Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19); Phosphoribosyl-ATP
DE   pyrophosphohydrolase (EC 3.6.1.31); Histidinol dehydrogenase
DE   (EC 1.1.1.23) (HDH)].
GN   HIS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 64385 / 1001;
RX   MEDLINE=98451822; PubMed=9778800;
RA   Navarro-Garcia F., Perez-Diaz R., Negredo A., Pla J., Nombela C.;
RT   "Cloning and sequence of a 3.835 kbp DNA fragment containing the HIS4
RT   gene and a fragment of a PEX5-like gene from Candida albicans.";
RL   Yeast 14:1147-1157(1998).
CC   -!- CATALYTIC ACTIVITY: 5-phosphoribosyl-AMP + H(2)O = 5-(5-phospho-D-
CC       ribosylaminoformimino)-1-(5-phospho-ribosyl) imidazole-4-
CC       carboxamide.
CC   -!- CATALYTIC ACTIVITY: 5-phosphoribosyl-ATP + H(2)O = 5-
CC       phosphoribosyl-AMP + diphosphate.
CC   -!- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2)O = L-histidine
CC       + 2 NADH.
CC   -!- COFACTOR: Binds 1 zinc ion (By similarity).
CC   -!- PATHWAY: Histidine biosynthesis; second step.
CC   -!- PATHWAY: Histidine biosynthesis; third step.
CC   -!- PATHWAY: Histidine biosynthesis; ninth (last) step.
CC   -!- SIMILARITY: In the C-terminal section; belongs to the histidinol
CC       dehydrogenase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ003115; CAA05871.1; -.
DR   InterPro; IPR001692; Histidinol_dh.
DR   InterPro; IPR002496; PRA-CH.
DR   InterPro; IPR008179; PRA-PH.
DR   InterPro; IPR008178; Pra_PH/CH.
DR   Pfam; PF00815; Histidinol_dh; 1.
DR   Pfam; PF01502; PRA-CH; 1.
DR   Pfam; PF01503; PRA-PH; 1.
DR   PRINTS; PR00083; HOLDHDRGNASE.
DR   ProDom; PD002680; Histidinol_dh; 1.
DR   ProDom; PD002610; PRA-CH; 1.
DR   ProDom; PD002611; Pra_PH/CH; 1.
DR   TIGRFAMs; TIGR00069; hisD; 1.
DR   PROSITE; PS00611; HISOL_DEHYDROGENASE; 1.
KW   Histidine biosynthesis; Multifunctional enzyme; Hydrolase;
KW   Oxidoreductase; NAD; Metal-binding; Zinc.
FT   DOMAIN        1    271       PHOSPHORIBOSYL-AMP CYCLOHYDROLASE.
FT   DOMAIN      272    360       PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE.
FT   DOMAIN      361    838       HISTIDINOL DEHYDROGENASE.
FT   ACT_SITE    729    729       BY SIMILARITY.
FT   ACT_SITE    730    730       BY SIMILARITY.
FT   METAL       660    660       ZINC (BY SIMILARITY).
FT   METAL       663    663       ZINC (BY SIMILARITY).
FT   METAL       764    764       ZINC (BY SIMILARITY).
FT   METAL       823    823       ZINC (BY SIMILARITY).
SQ   SEQUENCE   838 AA;  91835 MW;  BFA0E97E9CA62C03 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 838;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        482 GLSDEKLN 489
              ||||||||
Db        675 GLSDEKLN 682
```

```
RESULT 11
SYN_DROME
ID   SYN_DROME      STANDARD;      PRT;   980 AA.
AC   Q24546; Q24545;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Synapsin.
GN   SYN OR SYN1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS SYN-S AND SYN1-RT).
RC   STRAIN=Berlin; TISSUE=Head;
RX   MEDLINE=96209149; PubMed=8627354;
RA   Klagges B.R.E., Heimbeck G., Godenschwege T.A., Hofbauer A.,
RA   Pflugfelder G.O., Reifegerste R., Reisch D., Schaupp M., Buchner S.,
RA   Buchner E.;
RT   "Invertebrate synapsins: a single gene codes for several isoforms in
RT   Drosophila.";
RL   J. Neurosci. 16:3154-3165(1996).
CC   -!- SUBCELLULAR LOCATION: Synapse.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Syn1-RT;
CC         IsoId=Q24546-1; Sequence=Displayed;
CC       Name=Syn-S;
CC         IsoId=Q24546-2; Sequence=VSP_006324;
CC   -!- TISSUE SPECIFICITY: Widely expressed in the nervous system.
CC   -!- SIMILARITY: Belongs to the synapsin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95453; CAA64723.1; -.
DR   EMBL; X95453; CAA64722.1; -.
DR   HSSP; P17599; 1AUX.
DR   FlyBase; FBgn0004575; Syn.
DR   InterPro; IPR001359; Synapsin.
DR   Pfam; PF02078; Synapsin; 1.
DR   Pfam; PF02750; Synapsin_C; 1.
DR   PRINTS; PR01368; SYNAPSIN.
DR   PROSITE; PS00415; SYNAPSIN_1; FALSE_NEG.
DR   PROSITE; PS00416; SYNAPSIN_2; FALSE_NEG.
KW   Synapse; Neurone; Alternative splicing.
FT   DOMAIN        2      7       POLY-PRO.
FT   DOMAIN      668    680       POLY-SER.
FT   DOMAIN      692    695       POLY-PRO.
FT   DOMAIN      698    703       POLY-PRO.
FT   VARSPLIC    539    980       Missing (in isoform Syn-S).
FT                                /FTId=VSP_006324.
SQ   SEQUENCE   980 AA;  102799 MW;  E6037A2A4804F944 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 980;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        864 DTSGYDSN 871
              ||||||||
Db        823 DTSGYDSN 830


RESULT 12
TRM1_PYRWO
ID   TRM1_PYRWO     STANDARD;      PRT;    73 AA.
AC   P20300;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   N(2),N(2)-dimethylguanosine tRNA methyltransferase (EC 2.1.1.32)
DE   (tRNA(guanine-26,N(2)-N(2)) methyltransferase) (tRNA 2,2-
DE   dimethylguanosine-26 methyltransferase)
DE   (tRNA(m(2,2)G26)dimethyltransferase) (Fragment).
GN   TRM1.
OS   Pyrococcus woesei.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 3773;
RX   MEDLINE=90330536; PubMed=2165475;
RA   Zwickl P., Fabry S., Bogedain C., Haas A., Hensel R.;
RT   "Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic
RT   archaebacterium Pyrococcus woesei: characterization of the enzyme,
RT   cloning and sequencing of the gene, and expression in Escherichia
RT   coli.";
RL   J. Bacteriol. 172:4329-4338(1990).
RN   [2]
RP   SIMILARITY.
RX   MEDLINE=95206934; PubMed=7899076;
RA   Ouzounis C., Kyrpides N., Sander C.;
RT   "Novel protein families in archaean genomes.";
RL   Nucleic Acids Res. 23:565-570(1995).
CC   -!- FUNCTION: Dimethylates a single guanine residue at position 26 of
CC       a number of tRNAs using S-adenosyl-L-methionine as donor of the
CC       methyl groups (By similarity).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC       homocysteine + tRNA containing N(2)-methylguanine.
CC   -!- SIMILARITY: Belongs to the TRM1 family.
DR   PIR; S10650; S10650.
DR   HAMAP; MF_00290; -; 1.
DR   InterPro; IPR002905; TRM.
DR   Pfam; PF02005; TRM; 1.
KW   Transferase; Methyltransferase; tRNA processing.
FT   NON_TER      73     73
SQ   SEQUENCE   73 AA;  8094 MW;  A5413AEFDF00C00C CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1088 LLNVLKP 1094
              |||||||
Db         44 LLNVLKP 50


RESULT 13
GLH2_RAT
ID   GLH2_RAT       STANDARD;      PRT;   120 AA.
AC   P11963;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein hormones alpha chain 2 precursor (Follitropin alpha
DE   chain) (Follicle-stimulating hormone alpha chain) (FSH-alpha)
DE   (Lutropin alpha chain) (Luteinizing hormone alpha chain) (LSH-alpha)
DE   (Thyrotropin alpha chain) (Thyroid-stimulating hormone alpha chain)
DE   (TSH-alpha).
GN   CGA2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=82214055; PubMed=6177696;
RA   Godine J.E., Chin W.W., Habener J.F.;
RT   "Alpha subunit of rat pituitary glycoprotein hormones. Primary
RT   structure of the precursor determined from the nucleotide sequence of
RT   cloned cDNAs.";
RL   J. Biol. Chem. 257:8368-8371(1982).
CC   -!- SUBUNIT: Heterodimer of a common alpha chain and a unique beta
CC       chain which confers biological specificity to thyrotropin,
CC       lutropin, follitropin and gonadotropin.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the glycoprotein hormones alpha chain
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01253; CAA24566.1; -.
DR   HSSP; P01215; 1XUL.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR000476; Glyco_hormone.
DR   Pfam; PF00236; hormone6; 1.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   PRINTS; PR00274; GLYCOHORMONE.
DR   ProDom; PD002047; Glyco_hormone; 1.
DR   SMART; SM00067; GHA; 1.
DR   PROSITE; PS00779; GLYCO_HORMONE_ALPHA_1; 1.
DR   PROSITE; PS00780; GLYCO_HORMONE_ALPHA_2; 1.
DR   PROSITE; PS50277; GLYCO_HORMONE_ALPHA_3; 1.
KW   Hormone; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    120       GLYCOPROTEIN HORMONES ALPHA CHAIN 2.
FT   DISULFID     35     59       BY SIMILARITY.
FT   DISULFID     38     88       BY SIMILARITY.
FT   DISULFID     56    110       BY SIMILARITY.
FT   DISULFID     60    112       BY SIMILARITY.
FT   DISULFID     87    115       BY SIMILARITY.
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   CARBOHYD    106    106       N-LINKED (GLCNAC. . .) (PROBABLE).
SQ   SEQUENCE   120 AA;  13565 MW;  1BE324C29737B6D1 CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        120 PDGDFII 126
              |||||||
Db         26 PDGDFII 32


RESULT 14
GLHA_MASCO
ID   GLHA_MASCO     STANDARD;      PRT;   120 AA.
AC   Q9ERG4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein hormones alpha chain precursor (Follitropin alpha chain)
DE   (Follicle-stimulating hormone alpha chain) (FSH-alpha) (Lutropin alpha
DE   chain) (Luteinizing hormone alpha chain) (LSH-alpha) (Thyrotropin
DE   alpha chain) (Thyroid-stimulating hormone alpha chain) (TSH-alpha).
GN   CGA.
OS   Mastomys coucha (Southern multimammate mouse) (Praomys coucha).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC   Mastomys.
OX   NCBI_TaxID=35658;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MCC; TISSUE=Pituitary;
RA   Suzuki O.;
RT   "Praomys coucha glycoprotein hormone alpha-subunit.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: Heterodimer of a common alpha chain and a unique beta
CC       chain which confers biological specificity to thyrotropin,
CC       lutropin, follitropin and gonadotropin.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the glycoprotein hormones alpha chain
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF307149; AAG30277.1; -.
DR   HSSP; P01215; 1XUL.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR000476; Glyco_hormone.
DR   Pfam; PF00236; hormone6; 1.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   PRINTS; PR00274; GLYCOHORMONE.
DR   ProDom; PD002047; Glyco_hormone; 1.
DR   SMART; SM00067; GHA; 1.
DR   PROSITE; PS00779; GLYCO_HORMONE_ALPHA_1; 1.
DR   PROSITE; PS00780; GLYCO_HORMONE_ALPHA_2; 1.
DR   PROSITE; PS50277; GLYCO_HORMONE_ALPHA_3; 1.
KW   Hormone; Glycoprotein; Signal.
FT   SIGNAL        1     24       BY SIMILARITY.
FT   CHAIN        25    120       GLYCOPROTEIN HORMONES ALPHA CHAIN.
FT   DISULFID     35     59       BY SIMILARITY.
FT   DISULFID     38     88       BY SIMILARITY.
FT   DISULFID     56    110       BY SIMILARITY.
FT   DISULFID     60    112       BY SIMILARITY.
FT   DISULFID     87    115       BY SIMILARITY.
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    106    106       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   120 AA;  13473 MW;  532159F0E2C7A71F CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        120 PDGDFII 126
              |||||||
Db         26 PDGDFII 32


RESULT 15
GLHA_MERUN
ID   GLHA_MERUN     STANDARD;      PRT;   120 AA.
AC   Q9ERJ6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein hormones alpha chain precursor (Follitropin alpha chain)
DE   (Follicle-stimulating hormone alpha chain) (FSH-alpha) (Lutropin alpha
DE   chain) (Luteinizing hormone alpha chain) (LSH-alpha) (Thyrotropin
DE   alpha chain) (Thyroid-stimulating hormone alpha chain) (TSH-alpha).
GN   CGA.
OS   Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC   Meriones.
OX   NCBI_TaxID=10047;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pituitary;
RA   Suzuki O.;
```

```
RT   "Sequence of Mongolian gerbil glycoprotein alpha subunit cDNA.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: Heterodimer of a common alpha chain and a unique beta
CC       chain which confers biological specificity to thyrotropin,
CC       lutropin, follitropin and gonadotropin.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the glycoprotein hormones alpha chain
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF303351; AAG21399.1; -.
DR   HSSP; P01215; 1XUL.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR000476; Glyco_hormone.
DR   Pfam; PF00236; hormone6; 1.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   PRINTS; PR00274; GLYCOHORMONE.
DR   ProDom; PD002047; Glyco_hormone; 1.
DR   SMART; SM00067; GHA; 1.
DR   PROSITE; PS00779; GLYCO_HORMONE_ALPHA_1; 1.
DR   PROSITE; PS00780; GLYCO_HORMONE_ALPHA_2; 1.
DR   PROSITE; PS50277; GLYCO_HORMONE_ALPHA_3; 1.
KW   Hormone; Glycoprotein; Signal.
FT   SIGNAL        1     24       BY SIMILARITY.
FT   CHAIN        25    120       GLYCOPROTEIN HORMONES ALPHA CHAIN.
FT   DISULFID     35     59       BY SIMILARITY.
FT   DISULFID     38     88       BY SIMILARITY.
FT   DISULFID     56    110       BY SIMILARITY.
FT   DISULFID     60    112       BY SIMILARITY.
FT   DISULFID     87    115       BY SIMILARITY.
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    106    106       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   120 AA;  13473 MW;  251D3C573FC20205 CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        120 PDGDFII 126
              |||||||
Db         26 PDGDFII 32


Search completed: September 10, 2004, 05:07:16
Job time : 13.8033 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 50.8388 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 1250
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1093	87.4	1252	2	Q8KZM3	Q8kzm3 clostridium
2	1093	87.4	1255	2	Q9FAR6	Q9far6 clostridium
3	441	35.3	1251	2	Q9K395	Q9k395 clostridium
4	173	13.8	173	2	Q9KW90	Q9kw90 clostridium
5	150	12.0	367	2	Q45861	Q45861 clostridium
6	150	12.0	367	2	Q45862	Q45862 clostridium
7	114	9.1	173	2	Q9KW89	Q9kw89 clostridium
8	79	6.3	173	2	Q9KW88	Q9kw88 clostridium
9	76	6.1	173	2	Q9K345	Q9k345 clostridium
10	65	5.2	71	2	Q7WUH7	Q7wuh7 clostridium
11	65	5.2	71	2	Q7WRP2	Q7wrp2 clostridium
12	43	3.4	60	2	Q9R5H0	Q9r5h0 clostridium
13	42	3.4	52	2	Q9R5H1	Q9r5h1 clostridium
14	20	1.6	1268	2	Q45851	Q45851 clostridium
15	20	1.6	1278	2	Q57236	Q57236 clostridium
16	20	1.6	1280	2	Q9ZAJ5	Q9zaj5 clostridium
17	9	0.7	73	2	Q7WUH8	Q7wuh8 clostridium
18	9	0.7	361	2	Q45848	Q45848 clostridium
19	9	0.7	361	2	Q45846	Q45846 clostridium
20	9	0.7	375	16	Q8EMP0	Q8emp0 oceanobacil
21	9	0.7	1165	5	Q8I341	Q8i341 plasmodium
22	9	0.7	1291	2	Q9ZAJ8	Q9zaj8 clostridium
23	9	0.7	1291	2	Q93G71	Q93g71 clostridium
24	9	0.7	1291	2	Q933K0	Q933k0 clostridium
25	9	0.7	1291	2	Q08077	Q08077 clostridium
26	9	0.7	1291	2	Q8GR96	Q8gr96 clostridium
27	8	0.6	49	2	O87711	O87711 clostridium
28	8	0.6	82	16	Q82ZU5	Q82zu5 enterococcu
29	8	0.6	95	12	Q7TE86	Q7te86 dolichos ye
30	8	0.6	120	2	Q8KJT2	Q8kjt2 proteus vul
31	8	0.6	162	16	O32078	O32078 bacillus su
32	8	0.6	197	16	Q7U7I6	Q7u7i6 synechococc
33	8	0.6	209	11	Q9D8N8	Q9d8n8 mus musculu
34	8	0.6	306	16	Q8D8A0	Q8d8a0 vibrio vuln
35	8	0.6	306	16	Q87MY6	Q87my6 vibrio para
36	8	0.6	337	16	Q87UD6	Q87ud6 pseudomonas
37	8	0.6	342	10	Q9M9M8	Q9m9m8 arabidopsis
38	8	0.6	347	2	Q840G4	Q840g4 clostridium
39	8	0.6	377	16	Q7VYG2	Q7vyg2 bordetella
40	8	0.6	391	16	Q8RCZ3	Q8rcz3 thermoanaer
41	8	0.6	392	17	Q8TKZ6	Q8tkz6 methanosarc
42	8	0.6	392	17	Q8Q0P5	Q8q0p5 methanosarc
43	8	0.6	394	17	Q8TUI9	Q8tui9 methanosarc
44	8	0.6	397	2	Q9S523	Q9s523 escherichia
45	8	0.6	397	16	Q97GX2	Q97gx2 clostridium

ALIGNMENTS

RESULT 1
Q8KZM3
ID Q8KZM3 PRELIMINARY; PRT; 1252 AA.
AC Q8KZM3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 5262;
RA Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA Nakamura S., Karasawa T., Kozaki S.;
RT "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT complete cds.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB088207; BAC05434.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1252 AA; 143510 MW; 41B633BB744D3B41 CRC64;

Query Match 87.4%; Score 1093; DB 2; Length 1252;

```
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
```

```
RESULT 2
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match            87.4%;  Score 1093;  DB 2;  Length 1255;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 63

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 123

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 183

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 243

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        244 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 303

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 363

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 423

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 483

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 543

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        544 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 603

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 663

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 723

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 783

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 843

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 903

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 963

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1023

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1083

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1143

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1197


RESULT 3
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             35.3%;  Score 441;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  441;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEF 636
              |||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEF 636


RESULT 4
Q9KW90
ID   Q9KW90      PRELIMINARY;      PRT;   173 AA.
AC   Q9KW90;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Biwako;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35396;
RA   Kimura B.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35396;
RA   Kmura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Real-time, quantitative PCR assay of Clostriium botulinum type E in
RT   modified-atmosphere-packaged fish.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB040125; BAB07887.2; -.
DR   EMBL; AB040127; BAB07889.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19151 MW;  41781916CF06AB1F CRC64;

  Query Match             13.8%;  Score 173;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-170;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 60

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 120

Qy        580 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGI 632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGI 173


RESULT 5
Q45861
ID   Q45861      PRELIMINARY;      PRT;   367 AA.
AC   Q45861;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=type E;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70818; CAA50149.1; -.
DR   PIR; S48106; S48106.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42902 MW;  346A610C2FF70262 CRC64;

  Query Match             12.0%;  Score 150;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-146;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        817 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 261

Qy        877 DVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 DVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE 321

Qy        937 YTIINCMRDNNSGWKVSLNHNEIIWTLQDN 966
              |||||||||||||||||||||||||||||
Db        322 YTIINCMRDNNSGWKVSLNHNEIIWTLQDN 351


RESULT 6
Q45862
ID   Q45862      PRELIMINARY;      PRT;   367 AA.
AC   Q45862;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type E;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70815; CAA50146.1; -.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42854 MW;  0810595B3A865570 CRC64;

  Query Match             12.0%;  Score 150;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-146;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        817 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 DTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINING 261

Qy        877 DVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 DVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE 321

Qy        937 YTIINCMRDNNSGWKVSLNHNEIIWTLQDN 966
              |||||||||||||||||||||||||||||
Db        322 YTIINCMRDNNSGWKVSLNHNEIIWTLQDN 351


RESULT 7
Q9KW89
ID   Q9KW89      PRELIMINARY;      PRT;   173 AA.
AC   Q9KW89;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=5545;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
DR   EMBL; AB040126; BAB07888.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19137 MW;  4178024F5BC4BF1F CRC64;

  Query Match              9.1%;  Score 114;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-109;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 60

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFV 573
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFV 114


RESULT 8
Q9KW88
ID   Q9KW88      PRELIMINARY;      PRT;   173 AA.
AC   Q9KW88;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=164-1;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
DR   EMBL; AB040128; BAB07890.2; -.
DR   HSSP; P10845; 3BTA.
```

```
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19121 MW;  42F5822DE3B01F1F CRC64;

  Query Match              6.3%;  Score 79;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-73;
  Matches   79;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        495 DAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTF 554
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         36 DAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTF 95

Qy        555 FSSEFINNVNKPVQAALFV 573
              |||||||||||||||||||
Db         96 FSSEFINNVNKPVQAALFV 114


RESULT 9
Q9K345
ID   Q9K345      PRELIMINARY;      PRT;   173 AA.
AC   Q9K345; Q9KW91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Iwanai, and Tenno2;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
DR   EMBL; AB040123; BAB07885.2; -.
DR   EMBL; AB040124; BAB07886.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19061 MW;  5A89702DE3ACED16 CRC64;

  Query Match              6.1%;  Score 76;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-70;
  Matches   76;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        498 IPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSS 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 IPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSS 98

Qy        558 EFINNVNKPVQAALFV 573
              ||||||||||||||||
Db         99 EFINNVNKPVQAALFV 114


RESULT 10
Q7WUH7
ID   Q7WUH7      PRELIMINARY;      PRT;    71 AA.
AC   Q7WUH7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=21;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327859; AAQ16540.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      71     71
SQ   SEQUENCE   71 AA;  8391 MW;  A1B3F4F757EF6122 CRC64;

  Query Match              5.2%;  Score 65;  DB 2;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-58;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1079 YLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 1138
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 60

Qy       1139 DNLVR 1143
              |||||
Db         61 DNLVR 65


RESULT 11
Q7WRP2
ID   Q7WRP2      PRELIMINARY;      PRT;    71 AA.
AC   Q7WRP2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=109, and LCL155;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327860; AAQ16541.1; -.
DR   EMBL; AY327861; AAQ16542.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      71     71
SQ   SEQUENCE   71 AA;  8360 MW;  BDC3E11757EF6122 CRC64;

  Query Match              5.2%;  Score 65;  DB 2;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-58;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1079 YLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 1138
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 60

Qy       1139 DNLVR 1143
              |||||
Db         61 DNLVR 65


RESULT 12
Q9R5H0
ID   Q9R5H0      PRELIMINARY;      PRT;    60 AA.
```

```
AC   Q9R5H0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E neurotoxin, type E NT=16 kDa fragment (Fragment).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93000392; PubMed=1388670;
RA   Gimenez J.A., DasGupta B.R.;
RT   "Pepsin fragmentation of botulinum type E neurotoxin: isolation and
RT   characterization of 112, 48, 46, and 16 kD fragments.";
RL   J. Protein Chem. 11:255-264(1992).
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   InterPro; IPR002160; Kunitz_legume.
FT   NON_TER       1      1
FT   NON_TER      60     60
SQ   SEQUENCE   60 AA;  6865 MW;  9EA8C4CABFEAE766 CRC64;

  Query Match             3.4%;  Score 43;  DB 2;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-36;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1117 LANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHL 1159
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 LANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHL 43


RESULT 13
Q9R5H1
ID   Q9R5H1      PRELIMINARY;      PRT;    52 AA.
AC   Q9R5H1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E neurotoxin, type E NT=46 kDa fragment (Fragment).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93000392; PubMed=1388670;
RA   Gimenez J.A., DasGupta B.R.;
RT   "Pepsin fragmentation of botulinum type E neurotoxin: isolation and
RT   characterization of 112, 48, 46, and 16 kD fragments.";
RL   J. Protein Chem. 11:255-264(1992).
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
FT   NON_TER       1      1
FT   NON_TER      52     52
SQ   SEQUENCE   52 AA;  6122 MW;  38AE1305807188D1 CRC64;

  Query Match             3.4%;  Score 42;  DB 2;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-35;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        844 FKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNK 885
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 FKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNK 42


RESULT 14
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match             1.6%;  Score 20;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        702 INTQFNKRKEQMYQALQNQV 721


RESULT 15
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
```

RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELAND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).
DR EMBL; X81714; CAA57358.1; -.
DR EMBL; L35496; AAA23210.1; -.
DR EMBL; X70821; CAA50152.1; -.
DR EMBL; X99064; CAA67512.1; -.
DR PIR; S48110; S48110.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 1.6%; Score 20; DB 2; Length 1278;
Best Local Similarity 100.0%; Pred. No. 7.6e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 712 INTQFNKRKEQMYQALQNQV 731

Search completed: September 10, 2004, 05:12:50
Job time : 54.8388 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 82.4542 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 6511
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4033	61.9	804	5	ABG69083	Abg69083 Botulinum
2	2993.5	46.0	1059	3	AAY93309	Aay93309 A mangane
3	2993.5	46.0	1084	3	AAY93312	Aay93312 A mangane
4	2676	41.1	858	5	ABG69085	Abg69085 Botulinum
5	2313	35.5	1296	2	AAR95010	Aar95010 C. botuli
6	2305	35.4	449	3	AAY77137	Aay77137 Synthetic
7	2305	35.4	449	4	AAB04094	Aab04094 Botulism
8	2303	35.4	1295	5	AAU99339	Aau99339 Clostridi
9	2248.5	34.5	451	2	AAW68395	Aaw68395 Clostridi
10	2194	33.7	452	2	AAW68396	Aaw68396 Clostridi
11	2166	33.3	1291	2	AAW68392	Aaw68392 Clostridi
12	2157	33.1	419	4	AAB04095	Aab04095 Botulism
13	2125.5	32.6	1169	2	AAW56017	Aaw56017 Recombina
14	2120.5	32.6	1315	4	AAB61169	Aab61169 Clostridi
15	2112	32.4	422	5	ABG69071	Abg69071 Botulinum
16	2110	32.4	427	5	ABG69082	Abg69082 Botulinum
17	1916	29.4	382	4	AAB04100	Aab04100 Botulism
18	1792	27.5	1291	2	AAY05814	Aay05814 Non-toxic
19	1788	27.5	1067	3	AAY93307	Aay93307 A mangane
20	1788	27.5	1092	3	AAY93310	Aay93310 A mangane
21	1761	27.0	837	3	AAY77140	Aay77140 Native bo
22	1749.5	26.9	847	4	AAB04081	Aab04081 Botulism
23	1515	23.3	1070	3	AAY93308	Aay93308 A mangane
24	1515	23.3	1095	3	AAY93311	Aay93311 A mangane
25	1512.5	23.2	858	2	AAW56018	Aaw56018 Recombina
26	1504	23.1	1092	4	AAE07900	Aae07900 C. botuli
27	1497	23.0	852	5	ABG69077	Abg69077 Botulinum
28	1491	22.9	848	4	AAB04082	Aab04082 Botulism
29	1488.5	22.9	848	5	ABG69087	Abg69087 Botulinum
30	1482	22.8	1032	4	AAE07901	Aae07901 C. botuli
31	1457.5	22.4	408	4	AAB04101	Aab04101 Botulism
32	1442.5	22.2	1192	6	AAE35711	Aae35711 BoNT/F-Hc
33	1442.5	22.2	1192	6	AAE35710	Aae35710 BoNT/F-Hc
34	1423.5	21.9	979	6	AAE35713	Aae35713 BoNT/F-Hc
35	1422	21.8	887	4	AAE07892	Aae07892 Modified
36	1420	21.8	685	4	AAE07893	Aae07893 Modified
37	1420	21.8	862	4	AAE07890	Aae07890 Modified
38	1419.5	21.8	431	2	AAW09014	Aaw09014 Immunogen
39	1419.5	21.8	432	4	AAB04103	Aab04103 Botulism
40	1419.5	21.8	432	4	AAB04096	Aab04096 Botulism
41	1419.5	21.8	645	4	AAE07894	Aae07894 Modified
42	1419.5	21.8	645	6	AAE35692	Aae35692 DipT HN d
43	1419.5	21.8	657	6	AAE35693	Aae35693 BoNT/F-Hc
44	1419.5	21.8	657	6	AAE35694	Aae35694 BoNT/F-Hc
45	1402.5	21.5	432	3	AAY77138	Aay77138 Synthetic

ALIGNMENTS

RESULT 1
ABG69083
ID ABG69083 standard; protein; 804 AA.
XX
AC ABG69083;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #17.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98553.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 152-154; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

```
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 804 AA;

  Query Match             61.9%;  Score 4033;  DB 5;  Length 804;
  Best Local Similarity   97.3%;  Pred. No. 1.2e-209;
  Matches  782;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||||:|||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              | | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSIL 804
              |||||||||||||:|||:|||||
Db        781 KLREYDENVKTYLLNYIIQHGSIL 804


RESULT 2
AAY93309
```

```
ID   AAY93309 standard; protein; 1059 AA.
XX
AC   AAY93309;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 48-51; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC   be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype F. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a domain
CC   that translocates the SOD of the composition into the neuronal cell.
CC   After translocation, the linker is cleaved to release the SOD. The
CC   composition is useful for treating neuronal diseases caused or augmented
CC   by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC   disease, Huntington's disease and motor neurone diseases
XX
SQ   Sequence 1059 AA;

  Query Match             46.0%;  Score 2993.5;  DB 3;  Length 1059;
  Best Local Similarity   67.5%;  Pred. No. 2.5e-153;
  Matches  566;  Conservative  134;  Mismatches  120;  Indels   19;  Gaps    8;

Qy        425 ICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLS 484
              :|| :|| |||||||||:|||:::|||||||||||    |||| |:||:|||::|||: | :|
Db        225 LCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLNNNYRNNLDEVILDYNSETIPQIS 284

Qy        485 DEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTA 544
              :: ||  :|:|:|:|:|||||||:||:|:| :|||||||  |||||||| |::|||||||||
Db        285 NQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFFYLHAQKVPEGETNISLTSSIDTA 344

Qy        545 LLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVP 604
              | |: ::|||||||||| :|||| |||||:||| ||: |||||| |||| |||||||:|||
Db        345 LSEESQVYTFFSSEFINTINKPVHAALFISWINQVIRDFTTEATQKSTFDKIADISLVVP 404

Qy        605 YIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVI 664
              |:|||||||| || |||:| |||||||||| |||||||||||||||:|||:||||:|
Db        405 YVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELLIPTILVFTIKSFIGSSENKNKII 464
```

```
Qy        665 KAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYN 724
              |||||:| ||: ||||:||:|||||:|:|||||||||||||||||||:|:| :|| |||
Db        465 KAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTVIEYKYN 524

Qy        725 SYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLRE 784
              :||:|:| | ::|:|  |  |||:|||:|| ||:||:|||| |||||||| |::||||
Db        525 NYTSDERNRLESEYNINNIREELNKKVSLAMENIERFITESSIFYLMKLINEAKVSKLRE 584

Qy        785 YDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFF 844
              ||| || |||||| :| |||| | |||| :|  ||||||||:|||||:||||| |||| :
Db        585 YDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNNSIPFELSSYTNDKILILYFNKLY 644

Qy        845 KRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQND 904
              |:|| :|:|:|||:|:|::| ||| |||:|||||| | ||:|||||: | |||||:||:
Db        645 KKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNN 704

Qy        905 YIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQ 964
              |||: :|:||||||||||| | || ||:|||||||:|:|:|||||:|||:|:||||||
Db        705 DIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQ 763

Qy        965 DNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIH 1024
              | :| |||| |||     ||||||||||||||:|||:|::|||||||:||| |||:||
Db        764 DTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIH 823

Qy       1025 VSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDK 1084
              |||||||||| |: |||:||||| :|| || :|||:|||::||: :|||||||||||:|
Db        824 VSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNK 883

Qy       1085 EYYLLNVLKPNNFINRRTDSTLSINNIRS-----TILLANRLYSGIKVKIQRVNNSSTN- 1138
              |||||:|: :  | : : : |:||  |       |    |||:|::| |::  | ||:
Db        884 RYYLLNLLRTDKSITQNS-NFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRK--NGSTDI 940

Qy       1139 ---DNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKI---SSSGNRFNQVVVMNS 1192
                 || |||||  ||| |  :     |||| :    || ||:   |:| |   |::||:|
Db         941 SNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDS 999

Qy       1193 VG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQE 1249
              :| ||||||:|||| ||||||| :: :|||:||| ::| ||:||| ||:|||:||||||
Db       1000 IGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 1058
```

RESULT 3
AAY93312
ID AAY93312 standard; protein; 1084 AA.
XX
AC AAY93312;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;

XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 57-60; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide dismutase
CC (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC serotype F. The specification describes a composition for delivery of SOD
CC to neuronal cells. The composition comprises SOD linked, by a cleavable
CC linker, to a neuronal cell targeting component (NCTC). This component has
CC a domain that binds to a neuronal cell and a domain that translocates the
CC SOD of the composition into the neuronal cell. After translocation, the
CC linker is cleaved to release the SOD. The composition is useful for
CC treating neuronal diseases caused or augmented by oxidative stress, such
CC as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases
XX
SQ Sequence 1084 AA;

Query Match 46.0%; Score 2993.5; DB 3; Length 1084;
Best Local Similarity 67.5%; Pred. No. 2.6e-153;
Matches 566; Conservative 134; Mismatches 120; Indels 19; Gaps 8;

```
Qy        425 ICIEIINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLS 484
              :|| :|| |||||||||:|||:::|||||||||   |||| |:||:|||::|||: | :|
Db        250 LCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLNNNYRNNLDEVILDYNSETIPQIS 309

Qy        485 DEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTA 544
              :: ||  :|:|:|:|:|||||||:||:|:| :|||||| |||||||  |::||||||||
Db        310 NQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFFYLHAQKVPEGETNISLTSSIDTA 369

Qy        545 LLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVP 604
              | |: ::|||||||||| :|||| ||||:||| ||: |||||  |||| |||||||:|||
Db        370 LSEESQVYTFFSSEFINTINKPVHAALFISWINQVIRDFTTEATQKSTFDKIADISLVVP 429

Qy        605 YIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVI 664
              |:|||||||| || |||:| |||||||||| ||||||||||||||||:|||:||||:|
Db        430 YVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELLIPTILVFTIKSFIGSSENKNKII 489

Qy        665 KAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYN 724
              |||||:| ||: ||||:||:|||||:|:|||||||||||||||||||:|:| :|| |||
Db        490 KAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTVIEYKYN 549

Qy        725 SYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLRE 784
              :|| :|:| | ::|:|  |  |||:|||:|| ||:||:|||| |||||||| |::||||
Db        550 NYTSDERNRLESEYNINNIREELNKKVSLAMENIERFITESSIFYLMKLINEAKVSKLRE 609

Qy        785 YDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFF 844
              ||| || |||||| :| |||| | |||| :|  ||||||||:|||||:||||| |||| :
Db        610 YDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNNSIPFELSSYTNDKILILYFNKLY 669

Qy        845 KRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQND 904
              |:|| :|:|:|||:|:|::| ||| |||:|||||| | ||:|||||: | |||||:||:
Db        670 KKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNN 729

Qy        905 YIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQ 964
              |||: :|:||||||||||| | || ||:|||||||:|:|:|||||:|||:|:||||||
Db        730 DIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQ 788

Qy        965 DNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIH 1024
              | :| |||| |||     ||||||||||||||:|||:|::|||||||:||| |||:||
Db        789 DTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIH 848

Qy       1025 VSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDK 1084
              |||||||||| |: |||:||||| :|| || :|||:|||::||: :|||||||||||:|
Db        849 VSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNK 908
```

```
Qy       1085 EYYLLNVLKPNNFINRRTDSTLSINNIRS-----TILLANRLYSGIKVKIQRVNNSSTN- 1138
Db        909 RYYLLNLLRTDKSITQNS-NFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRK--NGSTDI 965

Qy       1139 ---DNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKI---SSSGNRFNQVVVMNS 1192
Db        966 SNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDS 1024

Qy       1193 VG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQE 1249
Db       1025 IGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 1083


RESULT 4
ABG69085
ID   ABG69085 standard; protein; 858 AA.
XX
AC   ABG69085;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #19.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98555.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 52; Page 157-159; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 858 AA;

  Query Match             41.1%;  Score 2676;  DB 5;  Length 858;
  Best Local Similarity   60.5%;  Pred. No. 2.8e-136;
  Matches  520;  Conservative  145;  Mismatches  162;  Indels   32;  Gaps   10;

Qy          1 MP-TINSFNYNDPVNNRTILYIK---PGGCQQPYKSFNIMKNIWIIPERNVIGTIPQDFL 56
Db          1 MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
Db         61 PPASLENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGN 120

Qy        117 DNTPDGDF-IINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNI-SLRNN---Y 171
Db        121 EHTPINEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVY 180

Qy        172 MPSNHGFGSIAIVTFSPEYSFRFKD------NSMNEFIQDPALTLMHELIHSLHGLYGAK 225
Db        181 DPSNDGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGAR 240

Qy        226 GITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
Db        241 GVTYKETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATR 300

Qy        286 LSKVQVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKF 343
Db        301 LSRVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKF 360

Qy        344 QVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRG 403
Db        361 KVKCRNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKG 420

Qy        404 LVKKIIRFCKNIVSVKGIR--KSI------------CIEINNGELFFVASENSYNDDNIN 449
Db        421 LVEKIVKFCKSVIPRKGTKNCKSVIPRKGTKAPPRLCIRVNNSELFFVASESSYNENDIN 480

Qy        450 TPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDI 509
Db        481 TPKEIDDTTNLNNNYRNNLDEVILDYNSQTIPQISNRTLNTLVQDNSYVPRYDSNGTSEI 540

Qy        510 EQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQA 569
Db        541 EEYDVVDFNVFFYLHAQKVPEGETNISLTSSIDTALLEESK-DIFFSSEFIDTINKPVNA 599

Qy        570 ALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLG 629
Db        600 ALFIDWISKVIRDFTTEATQKSTVDKIADISLIVPYVGLALNIIIEAEKGNFEEAFELLG 659

Qy        630 AGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNW 689
Db        660 VGILLEFVPELTIPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNW 719

Qy        690 MTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQ 749
Db        720 LTRINTQFNKRKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESEYNINNIEEELNK 779

Qy        750 KVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQ 809
Db        780 KVSLAMKNIERFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLLNYILDHRSILGEQTN 839

Qy        810 ELNSMVIDTLNNSIPFKLS 828
Db        840 ELSDLVTSTLNSSIPFELS 858


RESULT 5
AAR95010
ID   AAR95010 standard; protein; 1296 AA.
XX
AC   AAR95010;
XX
DT   09-JUL-1996  (first entry)
XX
DE   C. botulinum type A neurotoxin.
XX
```

KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.
XX
OS Clostridium botulinum.
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29244.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 4; Page 344-350; 434pp; English.
XX
CC Clostridium botulinum type A neurotoxin (AAR95010) is processed to form a
CC dimer composed of a light and a heavy chain. It is the product of the
CC type A neurotoxin gene (AAT29244). The 50 kDa C-terminal portion of the
CC heavy chain, or C fragment (see also AAR95008), was produced using a
CC synthetic gene (AAT29245) having codon usage altered to improve
CC expression in Escherichia coli. Fusion proteins of the type A toxin or C
CC fragment, with e.g. maltose binding protein or polyhistidine affinity tag
CC (see also AAR95008), are used to generate neutralising antitoxins and in
CC vaccine compsns
XX
SQ Sequence 1296 AA;

Query Match 35.5%; Score 2313; DB 2; Length 1296;
Best Local Similarity 40.0%; Pred. No. 1.9e-116;
Matches 536; Conservative 248; Mismatches 420; Indels 136; Gaps 44;

```
Qy          1 MPTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | ||| |  |  |  |:| |   ||:||||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||    |   ||||  || :| ||| :|| |||:| ||     ||:||  :  :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||       |:|:|:   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::           :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||       | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGE 433
              |   | :   ||     :  | | :: |:||  |                 :||::|| :
Db        409 NMNNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |        : : |  | || :: |  :|
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
               | |:  |       :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::|||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ::  |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | ||||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||| :    : :
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||     |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
               :: | | ::|:||:||:::  :| | | | |||   |   | :|||  ::| : |:: :
Db        867 TEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEV 926

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              : |:|:: |:||| |||:||| | | | ::|||||||||| :|||||||||: |||
Db        927 ILKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEII 984

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              |||||   | |:: | |     ||||||:||||||||:|| :||:|||| |||:| | ||
Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1021 GNIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNY 1079
              |||| |:||:||:   |  | ||| |:|||:|||||:| ||: ||:|: |: |||||||:|
Db       1045 GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDY 1104

Qy       1080 LLYDKEYYLLNVLKPNNFINRRTDSTLSINNI----------------RSTILLANRLYS 1123
              | ||| ||:||:  || :::        :||:                : | | : ||
Db       1105 LQYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK-EKTIKISSSG 1181
              | |  |::   |   ||:|| ||:|||| |  :| : |   | |   |    ::|   |
Db       1157 GTKFIIKKY-ASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVG 1215

Qy       1182 NRFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDN 1230
              |  :||||| |          | || ::||||:|| :||        :||| ||   :  :
Db       1216 N-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy       1231 TNSNGFFWNFISEEHGWQEK 1250
              : :  | || : || |:
Db       1275 SRTLGCSWEFIPVDDGWGER 1294
```

RESULT 6
AAY77137
ID AAY77137 standard; protein; 449 AA.
XX
AC AAY77137;
XX
DT 08-MAY-2000 (first entry)

```
XX
DE   Synthetic botulinum neurotoxin serotype E (BoNTE) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype E; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87215.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 26; Page 43-44; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 449 AA;

  Query Match             35.4%;  Score 2305;  DB 3;  Length 449;
  Best Local Similarity   96.7%;  Pred. No. 1.4e-116;
  Matches  434;  Conservative    8;  Mismatches    5;  Indels    2;  Gaps    1;

Qy        804 LGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 863
              :|||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 60

Qy        864 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRI 923
              ||||||||||||||||||||||||||||||:|:||||||||||||||||||||||||||
Db         61 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLTELNISQNDYIIYDNKYKNFSISFWVRI 120

Qy        924 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGI 983
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        121 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGI 180

Qy        984 SDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIG 1043
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        181 SDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIG 240

Qy       1044 IRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTD 1103
              |||||||||||||||||||:|||| ||||||||||||||||||||||||||||:|| |
Db        241 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKD 300

Qy       1104 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLY 1163
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLY 360

Qy       1164 ADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVAST 1221
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        361 ADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAST 420

Qy       1222 WYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||:||||| |||||||||||||
Db        421 WYYTHMRDHTNSNGCFWNFISEEHGWQEK 449


RESULT 7
AAB04094
ID   AAB04094 standard; protein; 449 AA.
XX
AC   AAB04094;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54488.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 7b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
```

```
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 449 AA;

  Query Match            35.4%;  Score 2305;  DB 4;  Length 449;
  Best Local Similarity  96.7%;  Pred. No. 1.4e-116;
  Matches  434;  Conservative    8;  Mismatches    5;  Indels    2;  Gaps    1;

Qy        804 LGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 863
              :|||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 60

Qy        864 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRI 923
              ||||||||||||||||||||||||||||||:|:|||||||||||||||||||||||||||
Db         61 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLTELNISQNDYIIYDNKYKNFSISFWVRI 120

Qy        924 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGI 983
              |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        121 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGI 180

Qy        984 SDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIG 1043
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        181 SDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIG 240

Qy       1044 IRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTD 1103
              ||||||||||||||||||:|||| |||||||||||||||||||||||:|| |
Db        241 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKD 300

Qy       1104 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLY 1163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLY 360

Qy       1164 ADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVAST 1221
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        361 ADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAST 420

Qy       1222 WYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||:||||| ||||||||||||||
Db        421 WYYTHMRDHTNSNGCFWNFISEEHGWQEK 449


RESULT 8
AAU99339
ID   AAU99339 standard; protein; 1295 AA.
XX
AC   AAU99339;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Clostridium botulinum botulism A toxin protein.
XX
KW   Botulism A toxin; neuroprotective; muscular; BoTOX; tetanus toxin; TeTOX;
KW   protease cleavage; toxicity; neurotoxin; poison; neural cell; toxin;
KW   post-translational modification; light chain; heavy chain;
KW   spastic condition; stabismus; bephlarospasm; hemifacial spasm;
KW   brain injury; spinal cord injury; stroke; multiple sclerosis;
KW   cerebral palsy; BoNT/A.
XX
OS   Clostridium botulinum.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .437
FT                   /note= "Light chain"
FT   Active-site     222
FT                   /note= "Forms an active site with Glu223, His226, Glu261,
FT                   Tyr365"
FT   Active-site     223
FT                   /note= "Forms an active site with His222, His226, Glu261,
FT                   Tyr365"
FT   Active-site     226
FT                   /note= "Forms an active site with His222, Glu223, Glu261,
FT                   Tyr365"
FT   Active-site     261
FT                   /note= "Forms an active site with His222, Glu223, His226,
FT                   Tyr365"
FT   Active-site     365
FT                   /note= "Forms an active site with His222, Glu223, His226,
FT                   Glu261"
FT   Region          437. .448
FT                   /note= "Linker region"
FT   Region          449. .1295
FT                   /note= "Heavy chain"
FT   Misc-difference 1161
FT                   /label= OTHER
FT                   /note= "OTHER = O, which is not defined in the
FT                   specification"
XX
PN   WO200244199-A2.
XX
PD   06-JUN-2002.
XX
PF   27-NOV-2001; 2001WO-US045059.
XX
PR   29-NOV-2000; 2000US-00726949.
XX
PA   (ALLR ) ALLERGAN SALES INC.
XX
PI   Lin W,  Aoki KR,  Steward LE;
XX
DR   WPI; 2002-557531/59.
XX
PT   Novel modified botulism toxin or tetanus toxin comprising a protease
PT   cleavage site, is useful for treating conditions benefited by neurotoxin
PT   activity.
XX
PS   Disclosure; Fig 1; 19pp; English.
XX
CC   The invention discloses modified botulism toxin (BoTOX) or tetanus toxin
CC   (TeTOX) which contain new protease cleavage sites. These sites are
CC   created in regions where the cleavage site is specifically susceptible to
CC   cleavage by a protease that is present in effective levels only in a
CC   tissue where toxicity is undesirable and where the cleavage site itself
CC   does not inactivate the toxin. BoTOX and TeTOX are neurotoxins which are
CC   highly potent and specific poisons of neural cells. The toxins are
CC   synthesised in vivo as single chains, which are not toxic, but become
CC   active when nicked, in a post-translational modification, to form
CC   separate light and heavy chains which are linked through a disulphide
CC   bond. The modified toxins are useful to treat conditions benefited by
CC   neurotoxin activity (e.g. spastic conditions, including stabismus,
CC   bephlarospasm and hemifacial spasm, brain injury, spinal cord injury,
CC   stroke, multiple sclerosis and cerebral palsy) by administering the
CC   toxins for their localised production. The advantage of the toxins are
CC   that they are deactivated in tissues where toxic activity is undesirable
CC   and activated at desired targets. The sequence presented is the
CC   Clostridium botulinum botulism A toxin (BoNT/A) protein
XX
SQ   Sequence 1295 AA;

  Query Match            35.4%;  Score 2303;  DB 5;  Length 1295;
  Best Local Similarity  39.9%;  Pred. No. 6.5e-116;
  Matches  534;  Conservative  247;  Mismatches  422;  Indels  136;  Gaps   44;

Qy          2 PTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DFL 56
              | :|  ||| ||||   | ||| |  |  |  |:| |   ||:||||:   | |:  |
Db          1 PFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLN 59

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              ||   |    |||| || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        117 DNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMPS 174
```

```
Db       120 -STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LT 175

Qy       175 NHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGAK 225

Db       176 RNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY--- 232

Qy       226 GITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADY 279

Db       233 GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKF 289

Qy       280 KKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 335

Db       290 KDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT 349

Qy       336 EFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 392

Db       350 EDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEIN 408

Qy       393 PRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGEL 434

Db       409 NMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDL 461

Qy       435 FFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSD 485

Db       462 FFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISI 516

Qy       486 EKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSID 542

Db       517 ENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVN 572

Qy       543 TALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIV 602

Db       573 EALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITII 632

Qy       603 VPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNK 662

Db       633 IPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVL 689

Qy       663 VIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESK 722

Db       690 TVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ 749

Qy       723 YNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKL 782

Db       750 YNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRL 807

Qy       783 REYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 841

Db       808 EDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFT 866

Qy       842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901

Db       867 EYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVI 926

Qy       902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961

Db       927 LKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIW 984

Qy       962 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1021

Db       985 DLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG 1044

Qy      1022 NIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db      1045 NIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYL 1104

Qy      1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNI---------------RSTILLANRLYSG 1124

Db      1105 QYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy      1125 IKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK-EKTIKISSSGN 1182

Db      1157 TKF-IKKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQVVVFKILSALEIPDVGN 1215

Qy      1183 RFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDNT 1231

Db      1216 -LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSS 1274

Qy      1232 NSNGFFWNFISEEHGWQEK 1250

Db      1275 RTLGCSWEFIPVDDGWGER 1293
```

RESULT 9
AAW68395
ID AAW68395 standard; protein; 451 AA.
XX
AC AAW68395;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain Belgua.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30584.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 41; Page 324-325; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC express BotE soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 451 AA;

Query Match 34.5%; Score 2248.5; DB 2; Length 451;
Best Local Similarity 99.8%; Pred. No. 1.6e-113;
Matches 424; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy       827 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 886
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          27 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 86

Qy         887 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 946
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          87 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 146

Qy         947 NSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 1006
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         147 NSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 206

Qy        1007 INGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNE 1066
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         207 INGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNE 266

Qy        1067 PNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIK 1126
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         267 PNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIK 326

Qy        1127 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQ 1186
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         327 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQ 386

Qy        1187 VVVMNSVG-CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEH 1245
                |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         387 VVVMNSVGNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEH 446

Qy        1246 GWQEK 1250
                |||||
Db         447 GWQEK 451
```

RESULT 10
AAW68396
ID AAW68396 standard; protein; 452 AA.
XX
AC AAW68396;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain NCTC 11219.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .21
FT   /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30585.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 41; Page 327-329; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC express BotE soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 452 AA;

```
  Query Match             33.7%;  Score 2194;  DB 2;  Length 452;
  Best Local Similarity   97.2%;  Pred. No. 1.4e-110;
  Matches  414;  Conservative    5;  Mismatches    5;  Indels    2;  Gaps    1;

Qy         827 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 886
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          27 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 86

Qy         887 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 946
                |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          87 QFGIDNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 146

Qy         947 NSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 1006
                ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db         147 NSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 206

Qy        1007 INGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNE 1066
                |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||:||
Db         207 INGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE 266

Qy        1067 PNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIK 1126
                || ||||||||||||||||||||||||||:|| ||||||||||||||||||||||||||||
Db         267 PNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIK 326

Qy        1127 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQ 1186
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         327 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQ 386

Qy        1187 VVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEE 1244
                ||||||||  ||||||||||||||||||||||||||||||:||||| ||||||||
Db         387 VVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEE 446

Qy        1245 HGWQEK 1250
                ||||||
Db         447 HGWQEK 452
```

RESULT 11
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX

```
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30579.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 291-296; 428pp; English.
XX
CC   This is the amino acid sequence of the type B toxin of Clostridium
CC   botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC   B toxin has been expressed as a histidine-tagged protein in Escherichia
CC   coli. The invention relates to C. botulinum recombinant toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of the recombinant proteins are E. coli, insect cells and
CC   yeast cells. The recombinant proteins are used as immunogens for the
CC   production of vaccines and antitoxins that are useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin.
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1291 AA;

  Query Match             33.3%;  Score 2166;  DB 2;  Length 1291;
  Best Local Similarity   37.3%;  Pred. No. 1.6e-108;
  Matches  490;  Conservative  267;  Mismatches  471;  Indels   84;  Gaps   32;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              ||  |||:|||||::|  |: ::|    |  ::||:|'|  |||||||  |  |:||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
                : : | |   ||||:|| :: :|: ||: : |:||||     |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   |  :|  | ||    :  :    :    |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   | :   |:         :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :       ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |     |  :| ||: |: ||   :|: | ||::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|         |: |||:: || | ||:||:  :::  :|||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :   |:   |    |: ||   |||     |||:::| :|||:| :|||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:  | :|  |||  ::::||: :  |   |  |        :  :| |:       |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
                :|  :| || :|       :::|||| | |||   |:|:||| ::|   || |:| |
```

```
Db        530 IFTDENTIFQYLYSQTFLLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::|  |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              ||||| |||||| :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751
              :||||   || ||:||  |  ||: ||: :|| |: :||: :    |   | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811
              : |::||: |:   |:|||||  :  : : || ::|  :|  ||:|| ::    |  | :
Db        765 NQAIHNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
              | |   |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |     :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINK 989
              ::|||||||||: ||||||:|:  | ||||| | :|  : : | |     ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | ||||||: | ::|:|||| |     | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS---- 1104
              ||: || :: |:  |  :  :  |||||||  |:|:|||||: |     |::|  : ||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLP 1161
                | |  |  |  :    || | |  |:| :|| | ||::||| | :|::|  :
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWR 1179

Qy       1162 LYADTATTNKEKTI---KISSSGNRFNQVVVMN-----SVGCTMNFKNN--NGNNIGLLG 1211
              :|       :|: :    || |   :| : :       :  | : || :  : : |||:|
Db       1180 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYHEQPTYSCQLLFKKDEESTDEIGLIG 1239

Qy       1212 FK------------ADTVVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249
                           |    | ||   ::    :   |  | || :: || |
Db       1240 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

```
RESULT 12
AAB04095
ID   AAB04095 standard; protein; 419 AA.
XX
AC   AAB04095;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
```

XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54489.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Fig 8; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 419 AA;

Query Match 33.1%; Score 2157; DB 4; Length 419;
Best Local Similarity 96.9%; Pred. No. 1.3e-108;
Matches 405; Conservative 7; Mismatches 4; Indels 2; Gaps 1;

Qy 835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894
Db 2 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 61

Qy 895 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 954
Db 62 LTELNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 121

Qy 955 NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDK 1014
Db 122 NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQ 181

Qy 1015 KSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKD 1074
Db 182 KSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKD 241

Qy 1075 FWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNN 1134
Db 242 FWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNN 301

Qy 1135 SSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
Db 302 SSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 361

Qy 1195 --CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
Db 362 NNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 419

RESULT 13
AAW56017
ID AAW56017 standard; protein; 1169 AA.
XX
AC AAW56017;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH728/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26289.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS Example 2; Page 91-94; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 1169 AA;

Query Match 32.6%; Score 2125.5; DB 2; Length 1169;
Best Local Similarity 39.2%; Pred. No. 2.2e-106;
Matches 466; Conservative 247; Mismatches 417; Indels 59; Gaps 26;

Qy 1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
Db 1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy 57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
Db 61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy 116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
Db 121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

```
Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  |  ||| |  : | |||   | :   |:          :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :      ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |      |  :| ||: |: ||   :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
              ::|  :::| | :|       |: |||:: || | ||:||:  :::  :|||| :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:  |    |: ||   |||     |||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  ||| ::::||: :  |   |  |         :  :| |:       |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
              :|  :| || :|  |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::| |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              |||||  |||||  :  | ::|::    |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751
              :||||   || ||:||  |  ||: ||: :|| |: :|: :    |   | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811
              : |::||: |:   |:||||| :  : : || ::|  :|  ||:|| ::   |  | :
Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
              | |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |    :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINK 989
                ::|||||||||: ||||||:|:  | |||||  | :|  : : | |     ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | ||||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS---- 1104
              ||: || :: |:  |  :  :  |||||| |:|:||||: |    |::|  : ||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYI 1150
                | |  |  |  :    || | |  |:| :|| | ||::||| | :|:
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYL 1169


RESULT 14
AAB61169
ID   AAB61169 standard; protein; 1315 AA.
XX
AC   AAB61169;
XX
DT   02-APR-2001  (first entry)
XX
DE   Clostridium tetani TeNT.
XX
KW   Clostridium tetani; TeNT; tetanus toxin; antibacterial; vaccine;
KW   TeNT fragment C; infection.
XX
OS   Clostridium tetani.
XX
PN   WO200100839-A1.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-GB002428.
XX
PR   25-JUN-1999;   99GB-00014861.
XX
PA   (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI   Fairweather NF,  Sinha K;
XX
DR   WPI; 2001-123014/13.
XX
PT   New polypeptide, useful for treating infections of Clostridium tetani,
PT   comprises tetanus toxin fragment with a mutation in a loop region,.
XX
PS   Disclosure; Page 39; 43pp; English.
XX
CC   The present sequence is given in a specification relating to a novel
CC   polypeptide comprising tetanus toxin (TeNT) fragment C or its immunogenic
CC   fragment, containing a mutation in a loop region. The mutation results in
CC   a reduction in the binding of TeNT fragment C or its immunogenic fragment
CC   to gangliosides and primary motoneurones, and/or a reduction in the
CC   ability of TeNT fragment C or its immunogenic fragment to undergo
CC   retrograde transport. The polypeptide is useful for treating, preventing
CC   and reducing the susceptibility to Clostridium tetani infection in a
CC   human or animal, and also for producing antibodies which recognise groups
CC   within TeNT polypeptides. Antibody produced against the polypeptide is
CC   also useful for treating Clostridium tetani infection
XX
SQ   Sequence 1315 AA;

  Query Match             32.6%;  Score 2120.5;  DB 4;  Length 1315;
  Best Local Similarity   35.6%;  Pred. No. 4.8e-106;
  Matches  486;  Conservative  262;  Mismatches  444;  Indels  175;  Gaps   38;

Qy          1 MP-TINSFNYNDPVNNRTILYIKPGGCQQ---FYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || |||:| |:|||||  ||: ::|  |:    :||:| |   |||:|||   || |:||
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

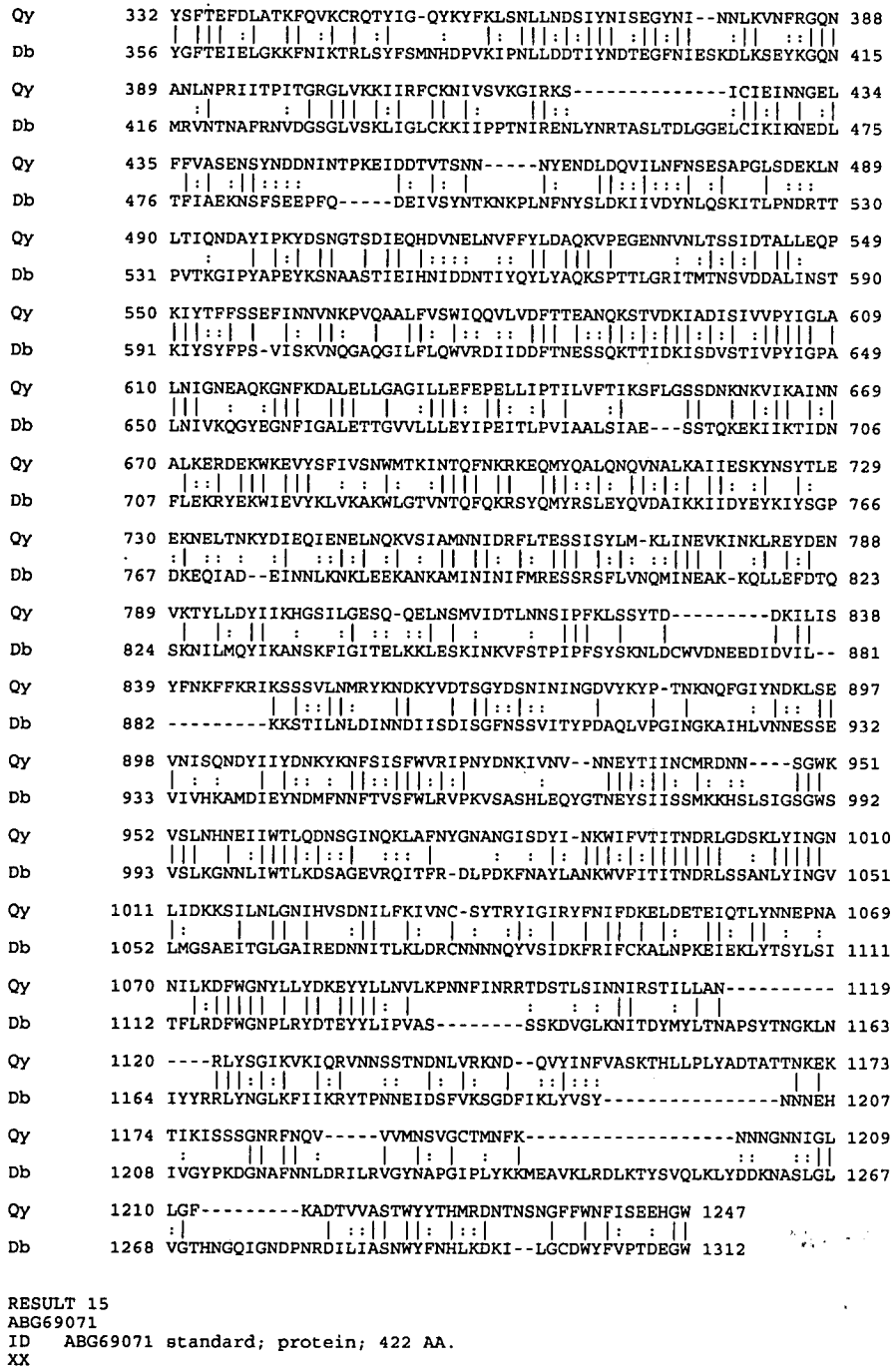
Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              ||:||  | | |||||||::| :||:||: : |:|||| :|::|  ||:::  | |||||
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        117 DNTPDGDFIINDASAVPIQFS------NG--SQSILLPNVIIMGAEPDLFETNSSNISLR 168
                :    |   | ::  : |:      :|  ::| :| |:|| |  | | :    | ||
Db        121 SYSLLDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLR 177

Qy        169 ---NNYMPSNHGFGSIAIVTFSPEYSFRFKDN----------SMNEFIQDPALTLMHELI 215
                 || |   |||||  : | |||   | ||            ::: ||||| ||||||
Db        178 VDNKNYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELI 236

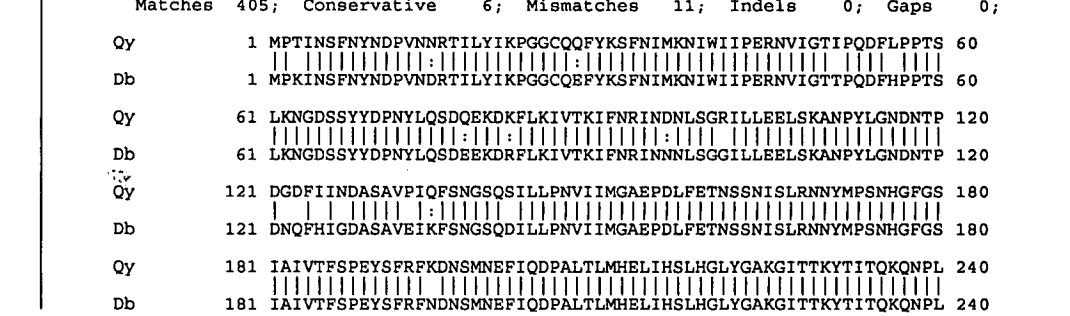
Qy        216 HSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275
              | |||||  : :::   |  ||   :  :    : ||  |||| | |:|:    ||:|
Db        237 HVLHGLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKT 295

Qy        276 LADYKKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-L 331
              | ||| ||:|||:|   |      :: || ::: ||  |||::| | || :||  ::   :
Db        296 LNDYKAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIM 355
```



RESULT 15
ABG69071
ID ABG69071 standard; protein; 422 AA.
XX
AC ABG69071;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #5.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98541.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 125-126; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 422 AA;

Query Match 32.4%; Score 2112; DB 5; Length 422;
Best Local Similarity 96.0%; Pred. No. 3.5e-106;
Matches 405; Conservative 6; Mismatches 11; Indels 0; Gaps 0;



```
Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IR 422
              ||
Db        421 IR 422
```

Search completed: September 10, 2004, 03:33:31
Job time : 86.4542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 26.5315 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 6511
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2313	35.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
2	2313	35.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
3	2313	35.5	1296	3	US-08-915-136-28	Sequence 28, Appl
4	2313	35.5	1296	4	US-09-084-517-28	Sequence 28, Appl
5	2125.5	32.6	1169	4	US-09-255-829-20	Sequence 20, Appl
6	2120.5	32.6	1315	4	US-08-913-880C-1	Sequence 1, Appli
7	1512.5	23.2	858	4	US-09-255-829-22	Sequence 22, Appl
8	1512.5	23.2	858	4	US-09-255-829-29	Sequence 29, Appl
9	1362	20.9	1013	4	US-09-255-829-18	Sequence 18, Appl
10	1355.5	20.8	907	4	US-09-255-829-16	Sequence 16, Appl
11	1351	20.7	871	4	US-09-255-829-8	Sequence 8, Appli
12	1343	20.6	871	4	US-09-255-829-2	Sequence 2, Appli
13	1343	20.6	871	4	US-09-255-829-26	Sequence 26, Appl
14	1343	20.6	875	4	US-09-255-829-10	Sequence 10, Appl
15	1341	20.6	873	4	US-09-255-829-6	Sequence 6, Appli
16	1341	20.6	894	4	US-09-255-829-4	Sequence 4, Appli
17	1340.5	20.6	878	4	US-09-255-829-12	Sequence 12, Appl
18	1340.5	20.6	953	4	US-09-255-829-14	Sequence 14, Appl
19	1284	19.7	858	4	US-08-913-880C-16	Sequence 16, Appl
20	1284	19.7	860	4	US-08-913-880C-15	Sequence 15, Appl
21	1284	19.7	862	4	US-08-913-880C-14	Sequence 14, Appl
22	1284	19.7	865	4	US-08-913-880C-13	Sequence 13, Appl
23	1284	19.7	866	4	US-08-913-880C-12	Sequence 12, Appl
24	1284	19.7	874	4	US-08-913-880C-11	Sequence 11, Appl
25	1284	19.7	875	4	US-08-913-880C-10	Sequence 10, Appl
26	1283	19.7	853	4	US-08-913-880C-17	Sequence 17, Appl
27	994	15.3	438	1	US-08-480-604A-23	Sequence 23, Appl
28	994	15.3	438	2	US-08-405-496A-23	Sequence 23, Appl
29	994	15.3	438	3	US-08-915-136-23	Sequence 23, Appl
30	994	15.3	438	4	US-09-084-517-23	Sequence 23, Appl
31	994	15.3	462	1	US-08-480-604A-26	Sequence 26, Appl
32	994	15.3	462	2	US-08-405-496A-26	Sequence 26, Appl
33	994	15.3	462	3	US-08-915-136-26	Sequence 26, Appl
34	994	15.3	462	4	US-09-084-517-26	Sequence 26, Appl
35	775	11.9	548	4	US-09-255-829-24	Sequence 24, Appl
36	561	8.6	618	1	US-08-668-381A-5	Sequence 5, Appli
37	560.5	8.6	452	1	US-07-618-312A-2	Sequence 2, Appli
38	560.5	8.6	452	1	US-08-280-228-2	Sequence 2, Appli
39	556.5	8.5	452	1	US-07-618-312A-4	Sequence 4, Appli
40	556.5	8.5	452	1	US-08-280-228-4	Sequence 4, Appli
41	527.5	8.1	452	1	US-08-110-786A-8	Sequence 8, Appli
42	368.5	5.7	207	3	US-08-446-114A-21	Sequence 21, Appl
43	310	4.8	2184	4	US-09-417-485D-6	Sequence 6, Appli
44	263.5	4.0	2710	1	US-08-480-604A-6	Sequence 6, Appli
45	263.5	4.0	2710	2	US-08-405-496A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             35.5%;  Score 2313;  DB 1;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 2e-140;
  Matches  536;  Conservative  248;  Mismatches  420;  Indels  136;  Gaps   44;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | ||| |  |  |  |:| |   ||:||||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||    |    ||||  || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
                :| | :  : | : :   :||      |::|:|  |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::          :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||        | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:|| || | :| || :||:  ||: ::| |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |    | :   ||      :  | | :: |:||  |               :||::|| :
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |        : :  |  | || :: |  :|
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
               | |:  |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::|||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ::  |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||   | ||||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||| :    : :
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
                                                                        
Db                                                                      
                                                                        
Qy                                                                      
                                                                        
Db                                                                      
                                                                        
Qy                                                                      
                                                                        
Db                                                                      
                                                                        
Qy                                                                      
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match           35.5%;  Score 2313;  DB 2;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 2e-140;
  Matches  536;  Conservative  248;  Mismatches  420;  Indels  136;  Gaps   44;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | ||| |  |  |  |:| |   ||:||||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||   |   |||| || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
                :| | :  : | : :   :||       |:|:|   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::           :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||       | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |  |:||: || | :| || :||:  ||: ::|  |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |    | :   ||       :  | | :: |:||  |               :||::|| :
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |        : : |  | || :: |  :|
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
               | |:  |       :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ::  |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | ||||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||  :    : :

Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||    |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              :: | | ::|:||:||:::  :| | | | |||    |    | :|||  ::| : |:: :
Db        867 TEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEV 926

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
                : |:|:: |:||| |||:||| | | | ::|||||||||| :|||||||||: |||
Db        927 ILKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEII 984

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              |||||   | |:: | |    ||||||:||||||||:|| :||:|||| |||:| | ||
Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1021 GNIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNY 1079
              |||| |:||:||:  |  | ||| |:|||:|||||:|  ||: ||:|: |: |||||:|
Db       1045 GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDY 1104

Qy       1080 LLYDKEYYLLNVLKPNNFINRRTDSTLSINNI----------------RSTILLANRLYS 1123
              | ||| ||:||:  || :::        :||:                 : | | : ||
Db       1105 LQYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK-EKTIKISSSG 1181
              | |  |::  |  ||:|| ||:|||| |  :| : |  |  |  |    ::|  |
Db       1157 GTKFIIKKY-ASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVG 1215

Qy       1182 NRFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDN 1230
              |  :||||| |          | || ::||||:|| :||        :||| ||   : :
Db       1216 N-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy       1231 TNSNGFFWNFISEEHGWQEK 1250
              : : |  | ||  : || |:
Db       1275 SRTLGCSWEFIPVDDGWGER 1294


RESULT 3
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
```

```
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             35.5%;  Score 2313;  DB 3;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 2e-140;
  Matches  536;  Conservative  248;  Mismatches  420;  Indels  136;  Gaps   44;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | ||| |  |  |  |:| |   ||:||||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||   |    |||| || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||     |::|:|  |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::          :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||          | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |    | :   ||       :  | | :: |:||  |               :||::|| :
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |         : :  |  | || :: |  :|
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              | |:  |       :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: ||  :  | : |::    ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              :: |:||| :|:||| ||| :||:||: |:||| :  ::: | :||:||  | ||||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||  :    : :
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  |: ||    |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              :: | | ::|:||:||:::  :| | | | |||   |   | :|||  ::| : |:: :
Db        867 TEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEV 926

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              : |:|:: |:||| |||:||| | | | ::|||||||||| :||||||||: |||
Db        927 ILKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEII 984

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              |||||   | |:: | |    ||||||:|||||||:|| :||:|||| |||:| | ||
Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1021 GNIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNY 1079
              |||| |:||:||:  |  | ||| |:|||:|||||:| ||: ||:|: |: |||||||:|
Db       1045 GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDY 1104

Qy       1080 LLYDKEYYLLNVLKPNNFINRRTDSTLSINNI----------------RSTILLANRLYS 1123
              | ||| ||:||:  || :::      :||:                 : | | : ||
Db       1105 LQYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK-EKTIKISSSG 1181
              | |  |::   |   ||:|| ||:|||| |  :| : |   |  |  |     ::|  |
Db       1157 GTKFIIKKY-ASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVG 1215

Qy       1182 NRFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDN 1230
              |  :||||| |          | || ::||||:|| :||         :||| ||   : :
Db       1216 N-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy       1231 TNSNGFFWNFISEEHGWQEK 1250
              : : |  | ||  : || |:
Db       1275 SRTLGCSWEFIPVDDGWGER 1294


RESULT 4
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             35.5%;  Score 2313;  DB 4;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 2e-140;
  Matches  536;  Conservative  248;  Mismatches  420;  Indels  136;  Gaps   44;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | ||| |  |  |  |:| |   ||:|||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||    |    ||||  || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
                :| | :  : | : : :   :||      |::|:|   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::            :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||         | |   ||    : |  :  ||  |||| |   | | | |: \
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |  |:||: || | :| || :||:  ||: ::| |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |    | :   ||       :  | | :: |:||  |               :||::|| :
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |        : : |  | || :: |  :|

Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
               | |:  |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              :: |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | ||||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||  :    : :
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||    |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              :: | | ::|:||:||:::  :| | | | |||   |   | :|||  ::| : |:: :
Db        867 TEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEV 926

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
                : |:|:: |:||| |||:||| | | | ::|||||||||| :||||||||||: |||
Db        927 ILKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEII 984

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              |||||   | |:: | |   ||||||:|||||||:|| :||:|||| |||:| | ||
Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1021 GNIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNY 1079
              |||| |:||:||:  |  | ||| |:|||:|||||:| ||: ||:|: |: |||||||:|
Db       1045 GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDY 1104

Qy       1080 LLYDKEYYLLNVLKPNNFINRRTDSTLSINNI----------------RSTILLANRLYS 1123
              | ||| ||:||:  || :::        :||:                : | | : ||
Db       1105 LQYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK-EKTIKISSSG 1181
              | |  |::   |   ||:|| ||:|||| |  :| : |   |  |   |   ::|  |
Db       1157 GTKFIIKKY-ASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVG 1215

Qy       1182 NRFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDN 1230
              |  :|||| |          | || ::||||:|| :||        :||| ||   :  :
Db       1216 N-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy       1231 TNSNGFFWNFISEEHGWQEK 1250
              : :  |  | ||  : || |:
Db       1275 SRTLGCSWEFIPVDDGWGER 1294


RESULT 5
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
```

```
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/255,829
;       FILING DATE:  23-FEB-1999
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB97/02273
;       FILING DATE:  22-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/782,893
;       FILING DATE:  27-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  ESMOND, ROBERT W.
;       REGISTRATION NUMBER:  32,893
;       REFERENCE/DOCKET NUMBER:  1581.0130002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1169 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-255-829-20

  Query Match             32.6%;  Score 2125.5;  DB 4;  Length 1169;
  Best Local Similarity   39.2%;  Pred. No. 2e-128;
  Matches  466;  Conservative  247;  Mismatches  417;  Indels   59;  Gaps   26;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || |||:|||||::|  |: ::|    |  ::||:| |   |||||||   |  |:||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
                : : | |   ||||:|| :: :|: ||: : |:||||     |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   |  :|  | ||    :  :    :     :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  |  ||| |  : | |||   | :   |:         :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :    :  ::   : :      ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |     |  :| ||: |: ||   :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|       |: |||:: || | ||:||:  :::  :||||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:   |    |: ||   |||     |||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  |||  ::::||: :  |   |  |         :  :| |:      |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
```

```
                .:|  :| || :|  |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::| |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              ||||| ||||||  :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751
               :||||   || ||:|| |  ||: ||: :|| |: :||: :    |  | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811
              : |::||: |:   |:|||| :  : : || ::|  :|  ||:|| ::    |  | :
Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
              | |   |  :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |    :|||| : :   |:: ::|| ||::: : :||:|||:||| | | :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINK 989
              ::|||||||||: ||||||:|:  | ||||| | :|  : : | |     ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | ||||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS---- 1104
              ||: || :: |:  |  :  :  ||||||| |:|:||||: |    |::|  : ||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

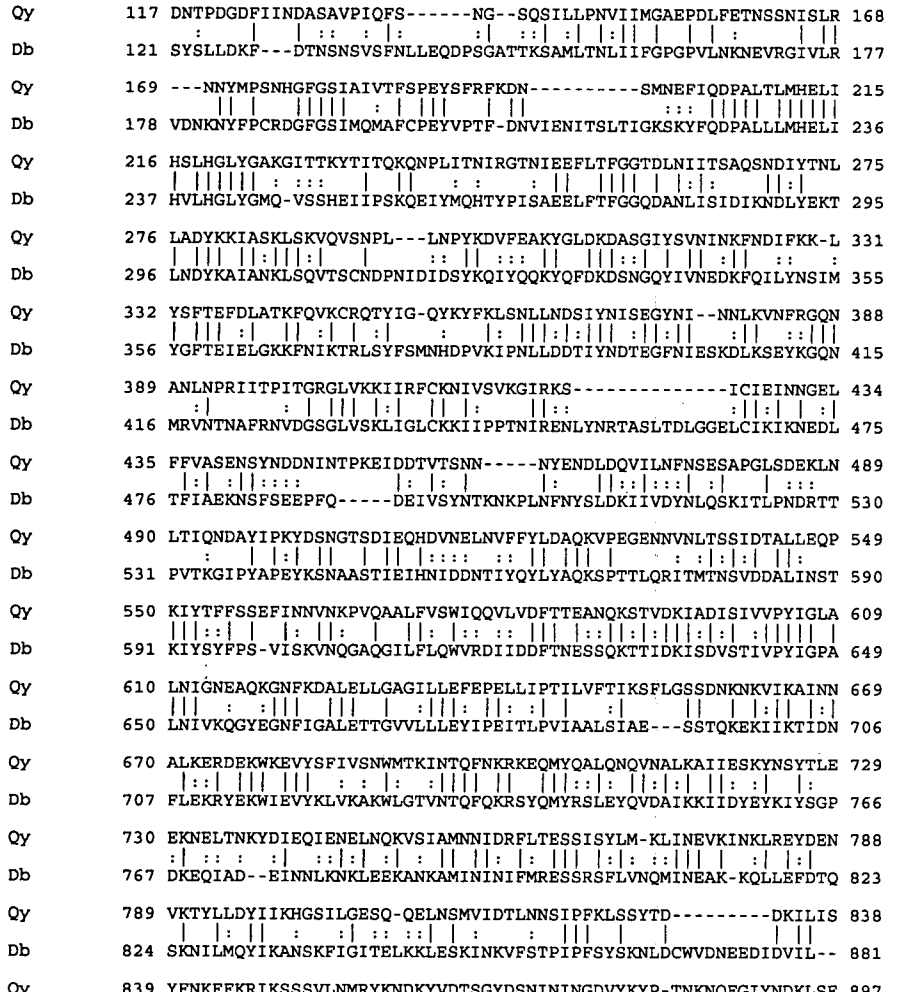
Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYI 1150
                | |  |  |  :    || | |  |:| :|| | ||::||| | :|:
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYL 1169

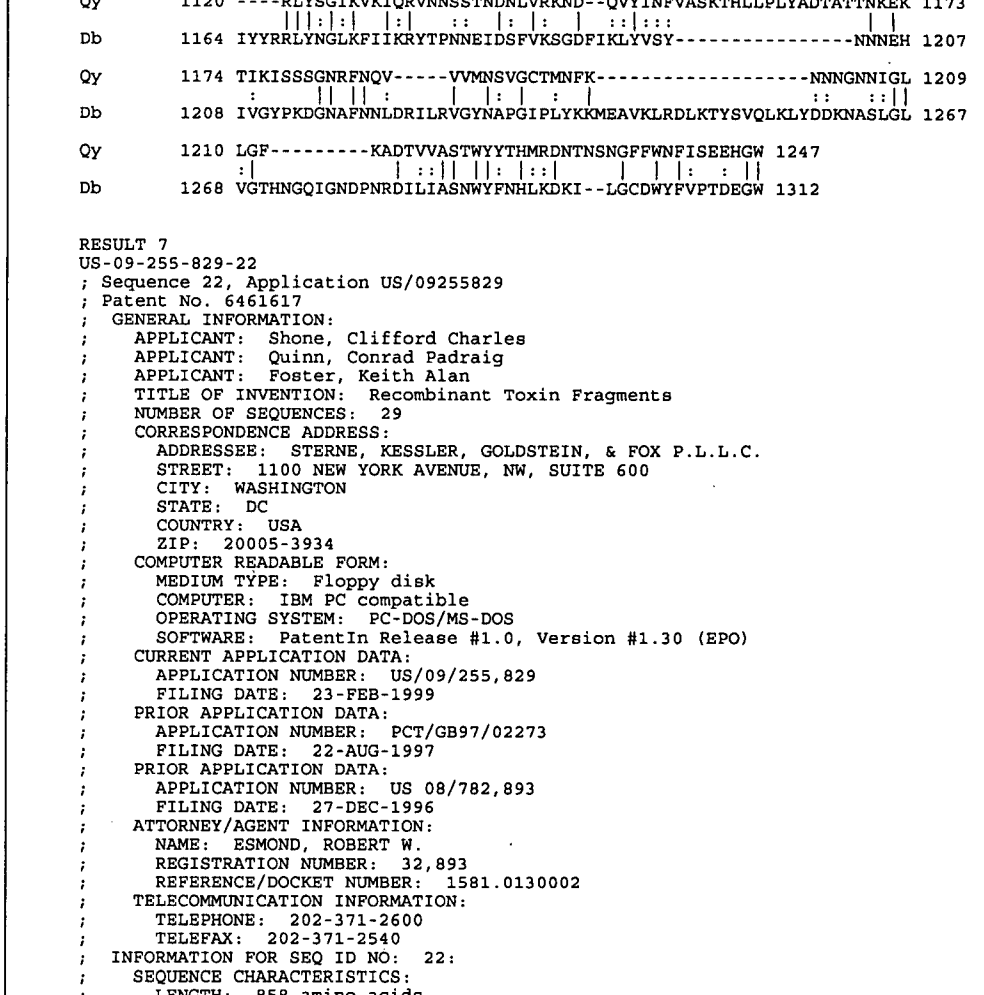

RESULT 6
US-08-913-880C-1
; Sequence 1, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
;  APPLICANT: MATSUDA, Morihiro
;  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
;  TITLE OF INVENTION:  VACCINE
;  FILE REFERENCE: 216-380P
;  CURRENT APPLICATION NUMBER: US/08/913,880C
;  CURRENT FILING DATE:  1997-11-24
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
;   LENGTH: 1315
;   TYPE: PRT
;   ORGANISM: Clostridium tetani
US-08-913-880C-1

  Query Match             32.6%;  Score 2120.5;  DB 4;  Length 1315;
  Best Local Similarity   35.6%;  Pred. No. 4.8e-128;
  Matches  486;  Conservative  262;  Mismatches  444;  Indels  175;  Gaps   38;

Qy          1 MP-TINSFNYNDPVNNRTILYIKPGGCQQ---FYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || |||:| |:|||||  ||: ::|  |:    :||:| |   |||:|||   || |:||
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

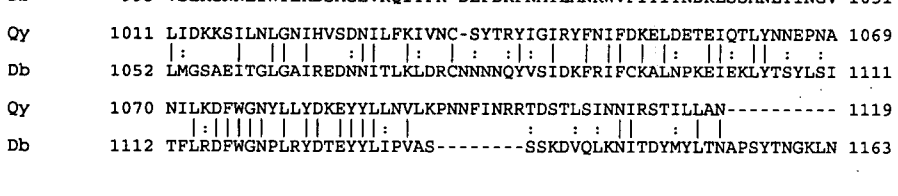
Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              ||:||  | | |||||||::| :||:|:: : |:|||| :|::|  ||:::  | |||||
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120
```

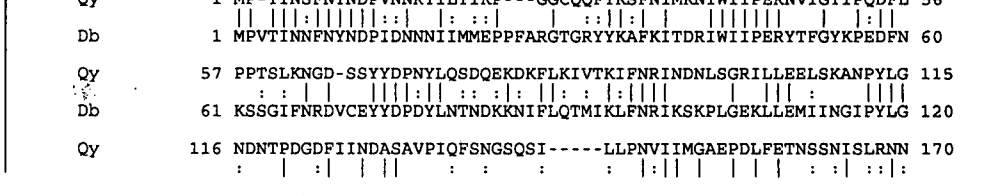




```
RESULT 7
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-22

  Query Match             23.2%;  Score 1512.5;  DB 4;  Length 858;
  Best Local Similarity   38.9%;  Pred. No. 3.6e-89;
  Matches  339;  Conservative  177;  Mismatches  311;  Indels   45;  Gaps   19;
```





```
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221

Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281

Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336

Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393

Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453

Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511

Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571

Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631

Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691

Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751

Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811

Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKF 843

Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKY 856


RESULT 8
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             23.2%;  Score 1512.5;  DB 4;  Length 858;
  Best Local Similarity   38.9%;  Pred. No. 3.6e-89;
  Matches  339;  Conservative  177;  Mismatches  311;  Indels   45;  Gaps   19;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56

Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115

Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170

Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221

Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281

Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336

Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393

Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453

Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511

Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571

Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631

Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
```

```
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751

Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811

Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKF 843

Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKY 856


RESULT 9
US-09-255-829-18
; Sequence 18, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1013 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-18

  Query Match             20.9%;  Score 1362;  DB 4;  Length 1013;
  Best Local Similarity   34.6%;  Pred. No. 2.1e-79;
  Matches  357;  Conservative  197;  Mismatches  370;  Indels  108;  Gaps   37;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55

Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115

Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173

Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224

Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278

Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334

Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391

Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPITG-RGLVKKIIRFC-KNIVSVK---------------GIRKSICIEINNGE 433

Db        409 NNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVNNWD 468

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484

Db        469 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 523

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541

Db        524 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 579

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601

Db        580 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 639

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661

Db        640 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 696

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721

Db        697 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 756

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781

Db        757 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 814

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        815 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 873

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPT-NKNQFGIYNDKLSEVN 899

Db        874 TEYIKSGLNSPGAAHYAQHDEAVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKD-D 932

Qy        900 ISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNE-----YTIINCMRDNNSGWKVSL 954

Db        933 PSQSANLLAEAKKLNDA-----QAPKVDNKF---NKEQQNAFYEILHLPNLNEE------ 978

Qy        955 NHNEIIWTLQDN 966

Db        979 QRNAFIQSLKDD 990


RESULT 10
US-09-255-829-16
; Sequence 16, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
```

```
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-16

  Query Match             20.8%;  Score 1355.5;  DB 4;  Length 907;
  Best Local Similarity   36.0%;  Pred. No. 4.8e-79;
  Matches  338;  Conservative  181;  Mismatches  324;  Indels   95;  Gaps   32;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||  | |||    | |  |:| |   ||:|||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||   |    ||||  || :| ||| :|| |||:| ||     ||:||  :  :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||       |:|:|   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::           :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||        | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |  :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
               ||  :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPITG-RGLVKKIIRFC-KNIVSVK---------------GIRKSICIEINNGE 433
              |    | :    || :     | : |:: |                |    :||::|| :

Db        409 NNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVNNWD 468

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |         : : |  | || :: |  :|
Db        469 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 523

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              | |:  |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        524 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 579

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::|||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        580 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 639

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        640 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 696

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              :: |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||   | ||||
Db        697 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 756

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||| :    : :
Db        757 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 814

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||     |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        815 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 873

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDV 878
              :: |    | |    :       ||| |:|: :  :
Db        874 TEYIKSRPQSKVKRQIF--------SGYQSDIDTHNRI 903


RESULT 11
US-09-255-829-8
; Sequence 8, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;
```

```
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-8

  Query Match           20.7%;  Score 1351;  DB 4;  Length 871;
  Best Local Similarity  36.9%;  Pred. No. 8.9e-79;
  Matches  334;  Conservative  176;  Mismatches  301;  Indels   94;  Gaps   32;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              || :|  ||| ||||  | |||    |  |  |:| |   ||:||||:   | |:  |
Db          1 MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||    |    |||| || :| ||| :|| |||:| ||      ||:||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||       |::|:|   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::           :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||         | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |    | :   ||       :  | | :: |:||  |                 :||::|| :
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |         : : |  | || :: |  :|
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              | |:  |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::|||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ::  |:||| :|:||| ||| :||:|: |:||| :  :::| :||:||  | ||||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||::|| : |:|||| :    : :
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||    |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFK 845


              :: |
Db        867 TEYIK 871


RESULT 12
US-09-255-829-2
; Sequence 2, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-2

  Query Match           20.6%;  Score 1343;  DB 4;  Length 871;
  Best Local Similarity  36.8%;  Pred. No. 2.9e-78;
  Matches  333;  Conservative  176;  Mismatches  302;  Indels   94;  Gaps   32;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | |||    |  |  |:| |   ||:||||:   | |:  |
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||    |    |||| || :| ||| :|| |||:| ||      ||:||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||       |::|:|   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::           :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||         | |   ||    : |  :  ||  |||| |   | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289
```

```
Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFK 845
Db        867 TEYIK 871
```

```
RESULT 13
US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-26

  Query Match             20.6%;  Score 1343;  DB 4;  Length 871;
  Best Local Similarity   36.8%;  Pred. No. 2.9e-78;
  Matches  333;  Conservative  176;  Mismatches  302;  Indels   94;  Gaps   32;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
```

```
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781

Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFK 845

Db        867 TEYIK 871


RESULT 14
US-09-255-829-10
; Sequence 10, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  875 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-10

  Query Match             20.6%;  Score 1343;  DB 4;  Length 875;
  Best Local Similarity   36.8%;  Pred. No. 2.9e-78;
  Matches  334;  Conservative  174;  Mismatches  305;  Indels   94;  Gaps   32;

Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55

Db          1 MQFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115

Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173

Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224

Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278

Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334

Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391

Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TG----------RGLVKKIIRFCKNIVSVKGIRKS-------ICIEINN 431

Db        409 NNMNFTKLKNFTGLFEFYKLLCVRGIITS-----KTKSLDKGYNKSADGALNDLCIKVNN 463

Qy        432 GELFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-G 482

Db        464 WDLFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPEN 518

Qy        483 LSDEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTS 539

Db        519 ISIENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTN 574

Qy        540 SIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADI 599

Db        575 SVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADI 634

Qy        600 SIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDN 659

Db        635 TIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---N 691

Qy        660 KNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAII 719

Db        692 KVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAII 751

Qy        720 ESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKI 779

Db        752 NYQYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGV 809

Qy        780 NKLREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILIS 838

Db        810 KRLEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLS 868

Qy        839 YFNKFFK 845

Db        869 TFTEYIK 875


RESULT 15
US-09-255-829-6
; Sequence 6, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
```

; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 873 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-6

Query Match 20.6%; Score 1341; DB 4; Length 873;
Best Local Similarity 36.8%; Pred. No. 3.9e-78;
Matches 333; Conservative 176; Mismatches 302; Indels 94; Gaps 32;

```
Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
              |  :|  ||| ||||   | |||    |  |  |:| |   ||:||||:   | |:  |
Db          3 MEFVNKQFNYKDPVNGVDIAYIKIPKYGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 61

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               ||    |    ||||  || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         62 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 121

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||      |::|:|   |:  ||  |    : |
Db        122 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 177

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| |:::          :|  |||:|| |||||: | ||
Db        178 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 235

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
               ||        | |   ||    : |  :  ||  |||| |   | | | |:
Db        236 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 291

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :
Db        292 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 351

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :
Db        352 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 410

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |    | :   ||      :  | | :: |:||  |               :||::|| :
Db        411 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 463

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||    ||: | |        : :  |  | || :: |  :|
Db        464 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 518

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              | |:  |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        519 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 574
```

```
Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::|||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|
Db        575 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 634

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        635 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 691

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              :: |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | ||||
Db        692 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 751

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|: : ::||: :: || ||:||| : |:||||  :    : :
Db        752 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 809

Qy        782 LREYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||    |:::|:  : |  | :||:  |||:|| | |:: |:| |
Db        810 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 868

Qy        841 NKFFK 845
               :: |
Db        869 TEYIK 873
```

Search completed: September 10, 2004, 03:47:00
Job time : 30.5315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 241.961 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 6511
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6511	100.0	1250	12	US-10-272-898-50	Sequence 50, Appl
2	6511	100.0	1250	15	US-10-354-774-50	Sequence 50, Appl
3	6511	100.0	1250	15	US-10-271-012-50	Sequence 50, Appl
4	6511	100.0	1250	16	US-10-729-122-50	Sequence 50, Appl
5	6511	100.0	1250	16	US-10-729-039-50	Sequence 50, Appl
6	6496.5	99.8	1251	15	US-10-452-024-5	Sequence 5, Appli
7	6496.5	99.8	1251	15	US-10-452-024-124	Sequence 124, App
8	6496	99.8	1252	15	US-10-452-024-125	Sequence 125, App
9	6496	99.8	1252	15	US-10-205-516-10	Sequence 10, Appl
10	6496	99.8	1255	15	US-10-452-024-126	Sequence 126, App
11	6484	99.6	1262	15	US-10-205-516-24	Sequence 24, Appl
12	6342	97.4	1252	12	US-10-272-898-52	Sequence 52, Appl
13	6342	97.4	1252	15	US-10-354-774-52	Sequence 52, Appl
14	6342	97.4	1252	15	US-10-271-012-52	Sequence 52, Appl
15	6342	97.4	1252	15	US-10-452-024-128	Sequence 128, App
16	6342	97.4	1252	16	US-10-729-122-52	Sequence 52, Appl
17	6342	97.4	1252	16	US-10-729-039-52	Sequence 52, Appl
18	6313.5	97.0	1251	15	US-10-452-024-129	Sequence 129, App
19	6171.5	94.8	1251	15	US-10-452-024-127	Sequence 127, App
20	4216	64.8	1278	15	US-10-452-024-152	Sequence 152, App
21	4216	64.8	1278	15	US-10-205-516-12	Sequence 12, Appl
22	4204	64.6	1288	15	US-10-205-516-26	Sequence 26, Appl
23	4145.5	63.7	1268	15	US-10-452-024-156	Sequence 156, App
24	4033	61.9	804	13	US-10-011-588-37	Sequence 37, Appl
25	4024.5	61.8	1280	15	US-10-452-024-162	Sequence 162, App
26	3986.5	61.2	1274	12	US-10-272-898-71	Sequence 71, Appl
27	3986.5	61.2	1274	15	US-10-354-774-71	Sequence 71, Appl
28	3986.5	61.2	1274	15	US-10-271-012-71	Sequence 71, Appl
29	3986.5	61.2	1274	15	US-10-452-024-6	Sequence 6, Appli
30	3986.5	61.2	1274	16	US-10-729-122-71	Sequence 71, Appl
31	3986.5	61.2	1274	16	US-10-729-039-71	Sequence 71, Appl
32	2676	41.1	858	13	US-10-011-588-41	Sequence 41, Appl
33	2322	35.7	1296	15	US-10-452-024-160	Sequence 160, App
34	2322	35.7	1296	15	US-10-205-516-2	Sequence 2, Appli
35	2321	35.6	1296	15	US-10-452-024-159	Sequence 159, App
36	2320	35.6	1306	15	US-10-205-516-16	Sequence 16, Appl
37	2316	35.6	1295	9	US-09-726-949A-1	Sequence 1, Appli
38	2316	35.6	1295	15	US-10-452-024-157	Sequence 157, App
39	2313	35.5	1296	12	US-10-272-898-28	Sequence 28, Appl
40	2313	35.5	1296	15	US-10-354-774-28	Sequence 28, Appl
41	2313	35.5	1296	15	US-10-271-012-28	Sequence 28, Appl
42	2313	35.5	1296	15	US-10-452-024-158	Sequence 158, App
43	2313	35.5	1296	16	US-10-729-122-28	Sequence 28, Appl
44	2313	35.5	1296	16	US-10-729-039-28	Sequence 28, Appl
45	2311	35.5	449	10	US-09-910-186A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-50
; Sequence 50, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-272-898-50

  Query Match             100.0%;  Score 6511;  DB 12;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 2
US-10-354-774-50
; Sequence 50, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-354-774-50

  Query Match             100.0%;  Score 6511;  DB 15;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
```

```
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 3
US-10-271-012-50
; Sequence 50, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-271-012-50

  Query Match           100.0%;  Score 6511;  DB 15;  Length 1250;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
```

```
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 4
```

```
US-10-729-122-50
; Sequence 50, Application US/10729122
; Publication No. US20040115215A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,122
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 50:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1250 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-729-122-50

  Query Match           100.0%;  Score 6511;  DB 16;  Length 1250;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 5
US-10-729-039-50
; Sequence 50, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-729-039-50

  Query Match            100.0%;  Score 6511;  DB 16;  Length 1250;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1250;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
```

```
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250

Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
```

RESULT 6
US-10-452-024-5
; Sequence 5, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-5

Query Match 99.8%; Score 6496.5; DB 15; Length 1251;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1249; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
```

```
Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-CTMNF 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNF 1200

Qy       1200 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1251


RESULT 7
US-10-452-024-124
; Sequence 124, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 124
;   LENGTH: 1251
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-124

  Query Match            99.8%;  Score 6496.5;  DB 15;  Length 1251;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches 1249;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-CTMNF 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNF 1200

Qy       1200 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1251


RESULT 8
US-10-452-024-125
; Sequence 125, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 125
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-125
```

```
  Query Match           99.8%;  Score 6496;  DB 15;  Length 1252;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1249;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1252
```

RESULT 9
US-10-205-516-10
; Sequence 10, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1252
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-10

```
  Query Match           99.8%;  Score 6496;  DB 15;  Length 1252;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1249;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600
Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1252
```

RESULT 10
US-10-452-024-126
; Sequence 126, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 126
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Clostridium butyricum
US-10-452-024-126

```
Query Match           99.8%;  Score 6496;  DB 15;  Length 1255;
Best Local Similarity 99.8%;  Pred. No. 0;
Matches 1249;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
Db          4 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 63
Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
Db         64 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 123
Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db        124 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 183
Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Db        184 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 243
Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db        244 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 303
Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db        304 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 363
Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db        364 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 423
Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db        424 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 483
Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db        484 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 543
Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        544 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 603
Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        604 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 663
Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
Db        664 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 723
Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db        724 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 783
Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
Db        784 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 843
Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db        844 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 903
Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db        904 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 963
Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
Db        964 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1023
Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
```

```
Db       1024 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1083

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1143

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db       1144 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1203

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1255


RESULT 11
US-10-205-516-24
; Sequence 24, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 1262
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-24

  Query Match             99.6%;  Score 6484;  DB 15;  Length 1262;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1249;  Conservative    0;  Mismatches    1;  Indels    6;  Gaps    2;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 I----RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFN 476
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRIEGRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFN 480

Qy        477 SESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVN 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        481 SESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVN 540

Qy        537 LTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKI 596
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        541 LTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKI 600

Qy        597 ADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGS 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGS 660

Qy        657 SDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALK 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALK 720

Qy        717 AIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINE 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINE 780

Qy        777 VKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKIL 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 VKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKIL 840

Qy        837 ISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS 900

Qy        897 EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNH 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNH 960

Qy        957 NEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKS 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 NEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKS 1020

Qy       1017 ILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFW 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFW 1080

Qy       1077 GNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS 1140

Qy       1137 TNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-- 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNN 1200

Qy       1195 CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1256


RESULT 12
US-10-272-898-52
; Sequence 52, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1252 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-272-898-52

  Query Match            97.4%;  Score 6342;  DB 12;  Length 1252;
  Best Local Similarity  97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   17;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              |||||||||||||||||| ||||||||||||||||||||||||||||||||  ||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              |||||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 13
US-10-354-774-52
; Sequence 52, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
```

```
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 1252 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-354-774-52

  Query Match            97.4%;  Score 6342;  DB 15;  Length 1252;
  Best Local Similarity  97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   17;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              |||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |   | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              |||||||||||||||||| ||||||||||||||||||||||||||||||||  ||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 14
US-10-271-012-52
; Sequence 52, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1252 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-271-012-52

  Query Match            97.4%;  Score 6342;  DB 15;  Length 1252;
```

```
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   17;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              |||||:|||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              |||||||||||||||||| ||||||||||||||||||||||||||||||||   ||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 15
US-10-452-024-128
; Sequence 128, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 128
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-128

  Query Match             97.4%;  Score 6342;  DB 15;  Length 1252;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   17;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
```

```
Qy       481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db       481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy       541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db       541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy       601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db       601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy       661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
Db       661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy       721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db       721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy       781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
Db       781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy       841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db       841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy       901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db       901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy       961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
Db       961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy      1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
Db      1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy      1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Db      1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy      1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
Db      1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy      1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
Db      1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252

Search completed: September 10, 2004, 04:58:05
Job time : 246.961 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 22.0831 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 6511
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6496.5	99.8	1251	2	JH0256	botulinum neurotox
2	6342	97.4	1252	2	S21178	botulinum neurotox
3	4145.5	63.7	1268	2	S33411	botulinum neurotox
4	3986.5	61.2	1274	2	I40813	neurotoxin type F
5	2322	35.7	1296	1	BTCLAB	bontoxilysin (EC 3
6	2264	34.8	1296	2	I40645	botulinum neurotox
7	2233	34.3	1297	2	S39791	neurotoxin - Clost
8	2181	33.5	1291	2	I40631	non-proteolytic bo
9	2176	33.4	1291	1	A48940	bontoxilysin (EC 3
10	2120.5	32.6	1315	1	BTCLTN	tentoxilysin (EC 3
11	1866	28.7	367	2	S48106	neurotoxin type E
12	1863	28.6	1276	2	S11455	botulinum neurotox
13	1852	28.4	1285	2	S70582	botulinum neurotox
14	1815	27.9	1291	2	S46431	botulinum neurotox
15	1815	27.9	1291	2	A49777	botulinum neurotox
16	1413.5	21.7	366	2	S48110	neurotoxin type F
17	1298	19.9	369	2	S48109	neurotoxin type F
18	766	11.8	1196	2	S46430	botulinum neurotox
19	765	11.7	1196	2	JQ1467	toxin, nontoxic co
20	746	11.5	1193	2	JC4901	nontoxic-nonhemagg
21	735	11.3	1162	2	A47708	progenitor toxin n
22	727.5	11.2	1162	2	I40817	botulinum toxin no
23	696.5	10.7	1193	2	S68218	botulinum neurotox
24	661.5	10.2	552	2	A49928	neurotoxin C1, BoN
25	655	10.1	1165	2	I40644	botulinum neurotox
26	437	6.7	122	2	A53878	type E neurotoxin
27	343.5	5.3	1711	2	T18429	hypothetical prote
28	342.5	5.3	1436	2	D71618	hypothetical prote
29	331.5	5.1	2269	2	T28677	rhoptry protein -
30	330.5	5.1	2401	2	T28676	rhoptry protein -
31	330.5	5.1	2485	1	H71621	serine/threonine-s
32	327.5	5.0	1127	2	T28317	ORF MSV156 hypothe
33	327.5	5.0	4688	2	F82885	hypothetical prote
34	325.5	5.0	3973	2	B71612	hypothetical prote
35	324.5	5.0	3724	2	T18427	hypothetical prote
36	321	4.9	5005	2	F82884	hypothetical prote
37	320.5	4.9	1516	2	E71619	RAD2 endonuclease
38	314.5	4.8	1844	2	D71612	hypothetical prote
39	314	4.8	1830	2	E82909	conserved hypothet
40	312	4.8	2380	2	E71604	hypothetical prote
41	310	4.8	2539	2	B71619	hypothetical prote
42	307	4.7	4550	2	T18440	hypothetical prote
43	303.5	4.7	4981	2	T18489	hypothetical prote
44	303	4.7	3844	2	T18402	asparagine/asparta
45	301	4.6	3394	2	T18501	hypothetical prote

ALIGNMENTS

RESULT 1
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.,
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxir
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release a
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 99.8%; Score 6496.5; DB 2; Length 1251;
Best Local Similarity 99.8%; Pred. No. 7.1e-258;
Matches 1249; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-CTMNF 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNF 1200

Qy       1200 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1251
```

RESULT 2
S21178

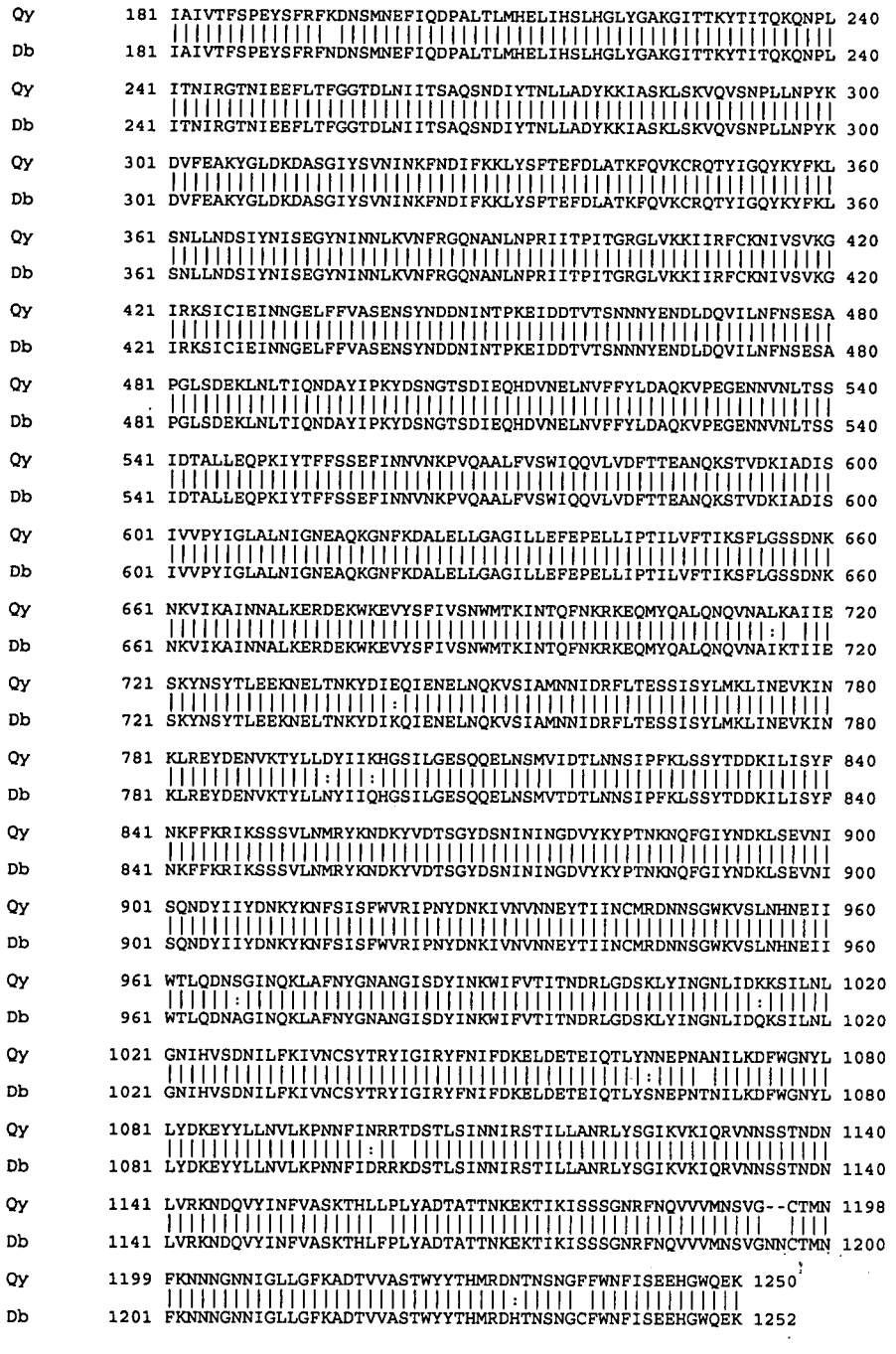
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin,
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match   97.4%; Score 6342; DB 2; Length 1252;
  Best Local Similarity 97.3%; Pred. No. 1.4e-251;
  Matches 1218; Conservative   15; Mismatches   17; Indels   2; Gaps   1;

```
Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||:|||||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

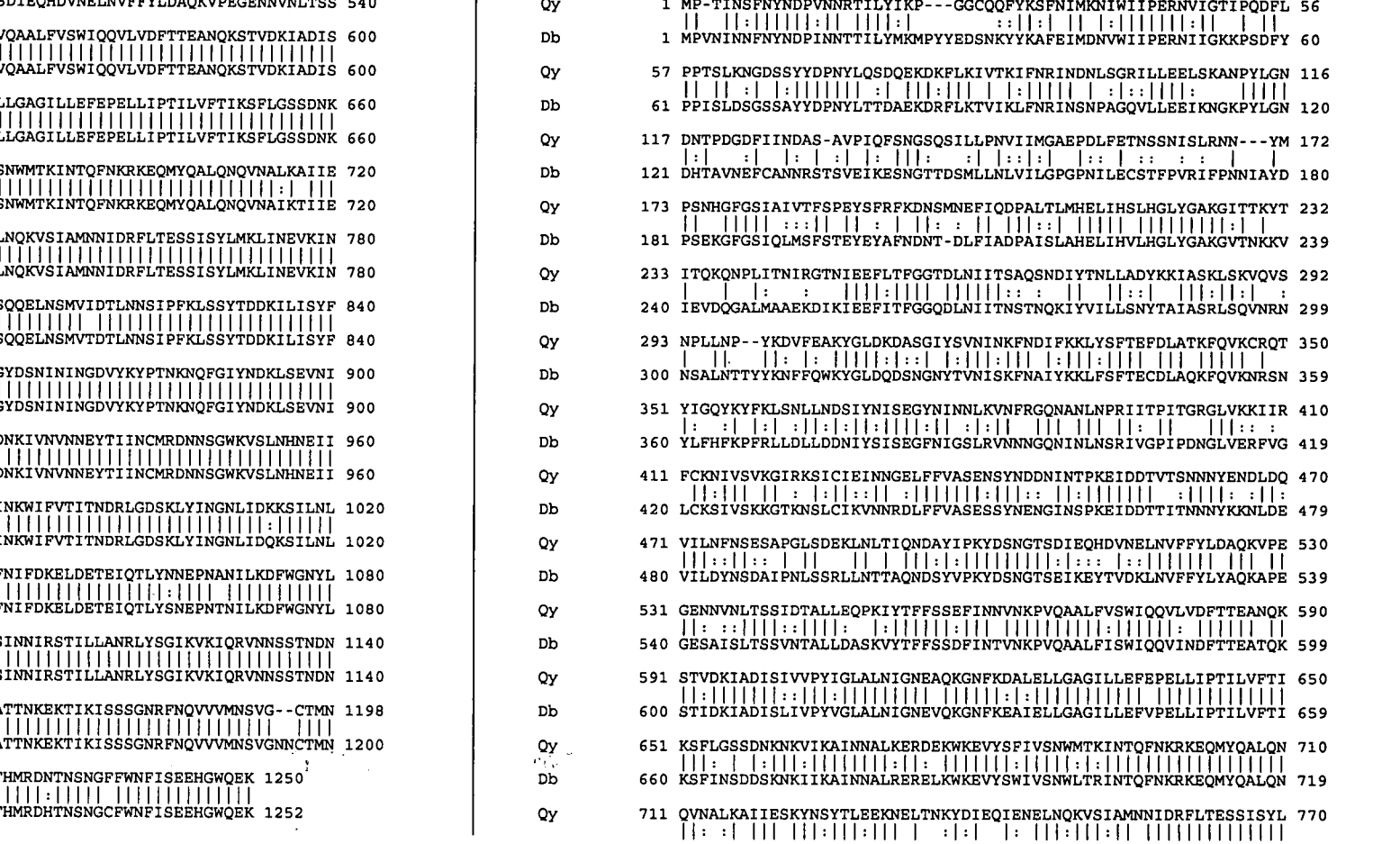
Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
```



RESULT 3
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 63.7%; Score 4145.5; DB 2; Length 1268;
Best Local Similarity 62.5%; Pred. No. 6.9e-162;
Matches 795; Conservative 221; Mismatches 229; Indels 27; Gaps 15;



```
Db        720 QVDGIKKIIEYKYNNYTLDEKNRLRAEYNIYSIKEELNKKVSLAMQNIDRFLTESSISYL 779

Qy        771 MKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQ-QELNSMVIDTLNNSIPFKLSS 829

Db        780 MKLINEAKINKLSEYDKRVNQYLLNYILENSSTLGTSSVPELNNLVSNTLNNSIPFELSE 839

Qy        830 YTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFG 889

Db        840 YTNDKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFG 899

Qy        890 IYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSG 949

Db        900 IYSSRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN-LKNLNNEYTIINCMRNNNSG 958

Qy        950 WKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYING 1009

Db        959 WKISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYING 1018

Qy       1010 NLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Db       1019 NLTDQKSILNLGNIHVDDNILFKIVGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR----STILLAN-RLYSG 1124

Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPNMSVTKNSD-ILNINRQRGIYSKTNIFSNARLYTG 1137

Qy       1125 IKVKIQRVNNSSTN--DNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIK---ISS 1179

Db       1138 VEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSE-YQLYADVSTSAVEKTIKLRRISN 1196

Qy       1180 SGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFF 1237

Db       1197 SNYNSNQMIIMDSIGDNCTMNFKTNNGNDIGLLGFHLNNLVASSWYYKNIRNNTRNNGCF 1256

Qy       1238 WNFISEEHGWQE 1249

Db       1257 WSFISKEHGWQE 1268


RESULT 4
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             61.2%;  Score 3986.5;  DB 2;  Length 1274;
  Best Local Similarity   60.3%;  Pred. No. 2.1e-155;
  Matches  772;  Conservative  224;  Mismatches  246;  Indels   39;  Gaps   17;

Qy          1 MP-TINSFNYNDPVNNRTILYIK---PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56

Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTNPSDFD 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116

Db         61 PPASLKNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYLGN 120

Qy        117 DNTPDGDFI-INDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRN------ 169

Db        121 DHTPIDEFSPVTRTTSVNIKLSTNVESSMLLNLLVLGAGPDIFE--SCCYPVRKLIDPDV 178

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFKD------NSMNEFIQDPALTLMHELIHSLHGLYG 223

Db        179 VYDPSNYGFGSINIVTFSPEYEYTFNDISGGHNSSTESFIADPAISLAHELIHALHGLYG 238

Qy        224 AKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIA 283

Db        239 ARGVTYEETIEVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIA 298

Qy        284 SKLSKVQVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLAT 341

Db        299 TRLSEVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTESDLAN 358

Qy        342 KFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITG 401

Db        359 KFKVKCRNTYFIKYEFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQSIKLNPKIIDSIPD 418

Qy        402 RGLVKKIIRFCKNIVSVKGIR--KSICIEINNGELFFVASENSYNDDNINTPKEIDDTVT 459

Db        419 KGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNSELFFVASESSYNENDINTPKEIDDTTN 478

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519

Db        479 LNNNYRNNLDEVILDYNSQTIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYDVVDFNV 538

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 579

Db        539 FFYLHAQKVPEGETNISLTSSIDTALLEESK-DIFFSSEFIDTINKPVNAALFIDWISKV 597

Qy        580 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPE 639

Db        598 IRDFTTEATQKSTVDKIADISLIVPYVGLALNIIIEAEKGNFEEAFELLGVGILLEFVPE 657

Qy        640 LLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNK 699

Db        658 LTIPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRINTQFNK 717

Qy        700 RKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNID 759

Db        718 RKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESEYNINNIEEELNKKVSLAMKNIE 777

Qy        760 RFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTL 819

Db        778 RFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLLNYILDHRSILGEQTNELSDLVTSTL 837

Qy        820 NNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVY 879

Db        838 NSSIPFELSSYTNDKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVY 897

Qy        880 KYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTI 939

Db        898 IYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTI 956

Qy        940 INCMRDNNSGWKVSL---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTIT 996

Db        957 INCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTIT 1016

Qy        997 NDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDE 1056

Db       1017 NNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDK 1076

Qy       1057 TEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR---- 1112

Db       1077 TEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS-GILNINQQRGVTE 1135
```

```
Qy       1113 STILLANRLYSGIKVKIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTN 1170
Db       1136 GSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSE 1193

Qy       1171 KEKTIKISSSGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMR 1228
Db       1194 KEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIR 1253

Qy       1229 DNTNSNGFFWNFISEEHGWQE 1249
Db       1254 RNTSSNGCFWSSISKENGWKE 1274
```

RESULT 5
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 35.7%; Score 2322; DB 1; Length 1296;
Best Local Similarity 40.0%; Pred. No. 2e-87;
Matches 536; Conservative 248; Mismatches 420; Indels 136; Gaps 43;

```
Qy          1 MPTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DF 55
Db          1 MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 LPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

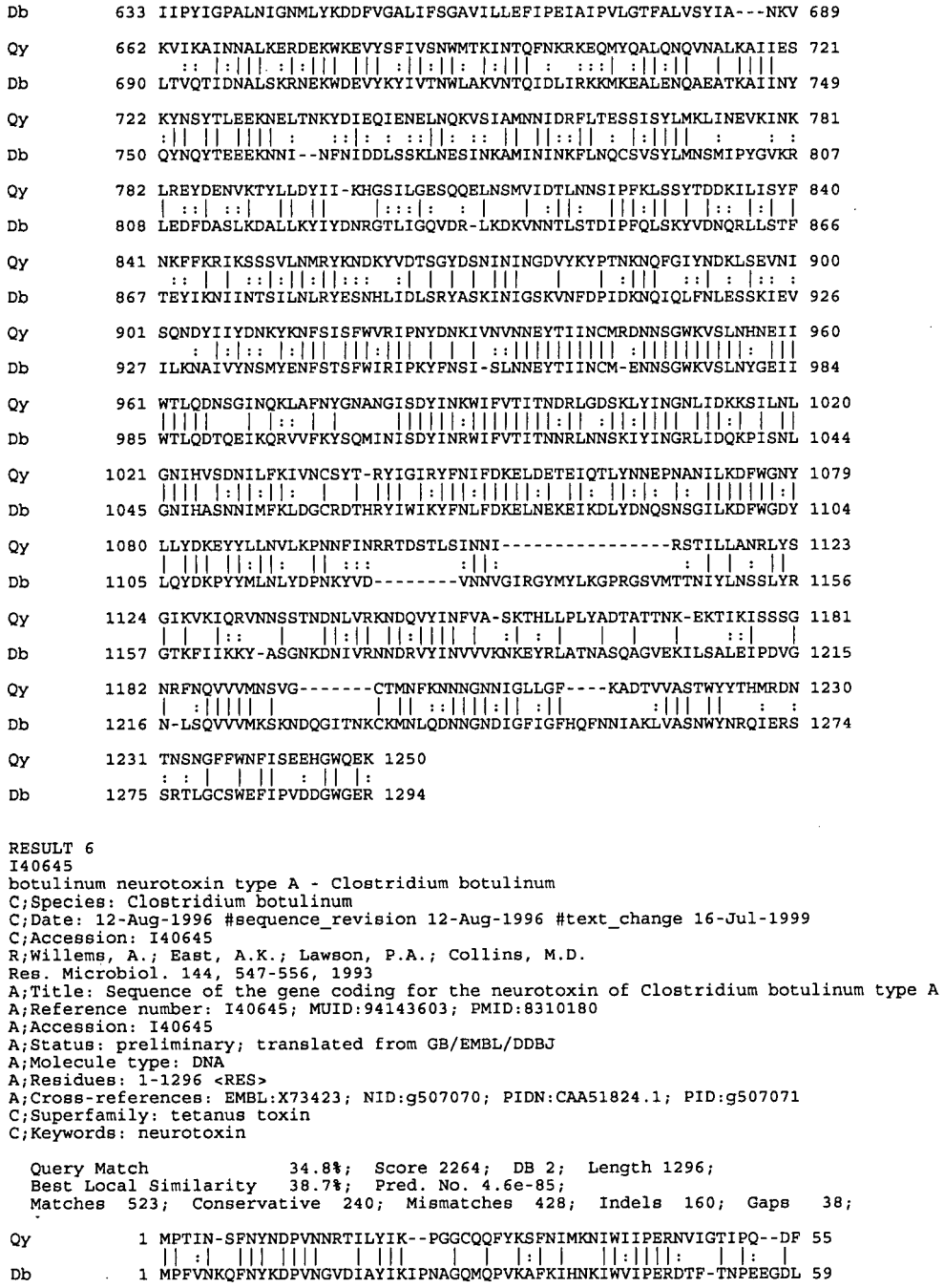
Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

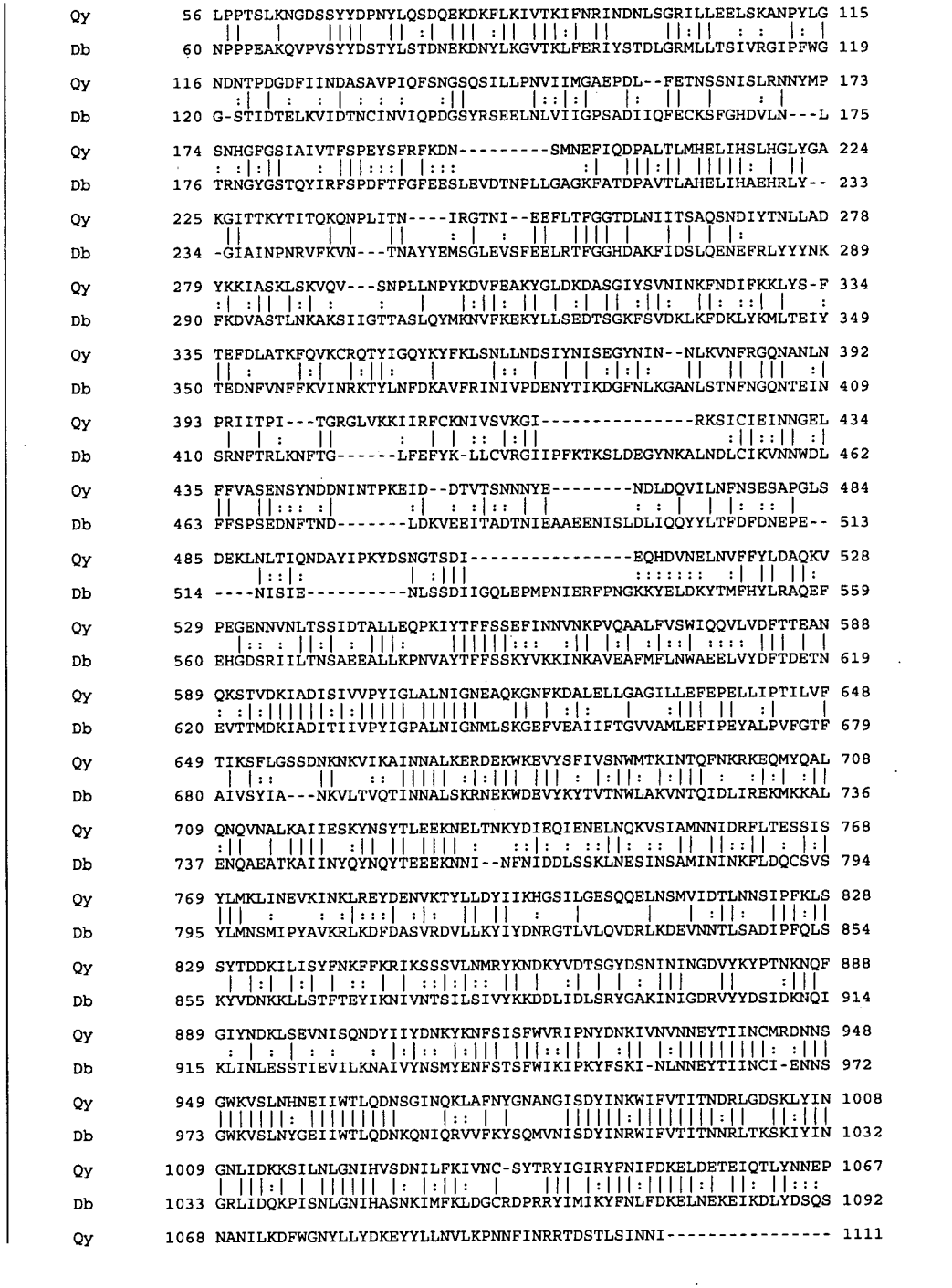
Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
```



RESULT 6
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 34.8%; Score 2264; DB 2; Length 1296;
Best Local Similarity 38.7%; Pred. No. 4.6e-85;
Matches 523; Conservative 240; Mismatches 428; Indels 160; Gaps 38;



```
                |: ||||||||||| ||| ||:||:  || :::        :|||
Db       1093 NSGILKDFWGNYLQYDKPYYMLNLFDPNKYVD--------VNNIGIRGYMYLKGPRGSVV 1144

Qy       1112 RSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTN 1170
               : | | : || | |  |::   |   ||:|| ||:|||| |  :| : |   |  |
Db       1145 TTNIYLNSTLYEGTKFIIKKY-ASGNEDNIVRNNDRVYINVVVKNKEYRLATNASQAGVE 1203

Qy       1171 K-EKTIKISSSGNRFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVV 1218
                |    ::|    ||  :|||||  |          | || ::||||:|| :||        :|
Db       1204 KILSALEIPDVGN-LSQVVVMKSKDDQGIRNKCKMNLQDNNGNDIGFIGFHLYDNIAKLV 1262

Qy       1219 ASTWYYTHMRDNTNSNGFFWNFISEEHGWQE 1249
                || ||   :   : : |  | ||  : || |
Db       1263 ASNWYNRQVGKASRTFGCSWEFIPVDDGWGE 1293


RESULT 7
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             34.3%;  Score 2233;  DB 2;  Length 1297;
  Best Local Similarity   38.7%;  Pred. No. 8.4e-84;
  Matches  511;  Conservative  248;  Mismatches  466;  Indels   96;  Gaps   34;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
                ||  | :|||||:||  |: ::|    |  :||:| |:  |||:|||   |  |  |
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
               | : : |   |||| ||::| ||||||| : |:|||||   ||: ||: :  | ||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQ---FSNGSQSI---LLPNVIIMGAEPDLFETNSSNISLRN 169
               | :||   |  | |: | |       |::     |: |:|| |  | | : | ::  : |
Db        121 NASTPPDKFAANVAN-VSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMN 178

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRF------KDNSM---NEFIQDPALTLMHELIHSLHG 220
               : | : |||:  ::  | |      |      || |:     :  ||||||||||| |||
Db        179 GHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHG 238

Qy        221 LYGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTN 274
               |||  |     ||  ||     |  :|:          ||  |||| | ::|: :   :||
Db        239 LYGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNK 291

Qy        275 LLADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL 331
               | ::: ||::|    |  | |   :: || ::: ||    :| :| |||: :||: ::| |
Db        292 ALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKAL 351

Qy        332 -YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQ 387
               : ||| :|| :: :| | :|  :|     |   ||:::||  :||:||   |||  | ||
Db        352 MFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQ 411

Qy        388 NANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDN 447
               |  :|       |:   ||   |  || ::     :   || :|| :|||:|:::|:: |
Db        412 NKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDL 471

Qy        448 INTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKY 501
                      :|:   | ||  ||:  :||:||: :  |   | :|     |  :|  || |
Db        472 AKA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVY 526

Qy        502 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 561
                    | :::  |:  ::| || ||  |    |: ||:|:: ||    |:|||||:  :
Db        527 IKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVE 584

Qy        562 NVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 621
                 |  | |:|||:|:: |: |||:|: ||||:||::|:||::|||| |||:|||  | ||
Db        585 KANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENF 644

Qy        622 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 681
               |:| |: || ||:|| |||::|  :  ||::|::|   ||  :|  |:||||:||:|| ::
Db        645 KNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDM 701

Qy        682 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIE 741
                |  ||| |:: :||||   ||:|| || ||  |:: ||| :|| |: |:|  :    |
Db        702 YGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFN 759

Qy        742 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHG 801
                |: :||| :::|:|||| |: : ||||||  :  : : ||:::|:|:|  ||:||  :
Db        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819

Qy        802 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 861
                | :    | | |   | :|||| || || | |||  || :   | |:::|:: |:  :
Db        820 LYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGR 879

Qy        862 YVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWV 921
                :|:||| : :|:  ||         || : | : | :   |: :::||: : ||||:|||
Db        880 LIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWV 939

Qy        922 RIPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGN 979
                | | |:|  :   : ||||||:|:: |:||||||:  | ||||| | :  :: : | |
Db        940 RTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSI 998

Qy        980 ANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCS-Y 1038
                : |||||||  :|||||||:: :||||:|   : |||  |: |::| ||::||:
Db        999 KDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDT 1058

Qy       1039 TRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFI 1098
                |::: |: |||| :||: ||: :||  : : | |||||||  | || :||| |    | :|
Db       1059 TKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYI 1118

Qy       1099 ----NRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN- 1151
                     |    |  :  :    || |::  |:: :||    |||:||: | :|:|
Db       1119 KYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI 1178

Qy       1152 ------------FVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMN-SVGCTMN 1198
                         | ||      |:   |  | : |         :: :    |  : |  :
Db       1179 DNISDESYRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD 1236

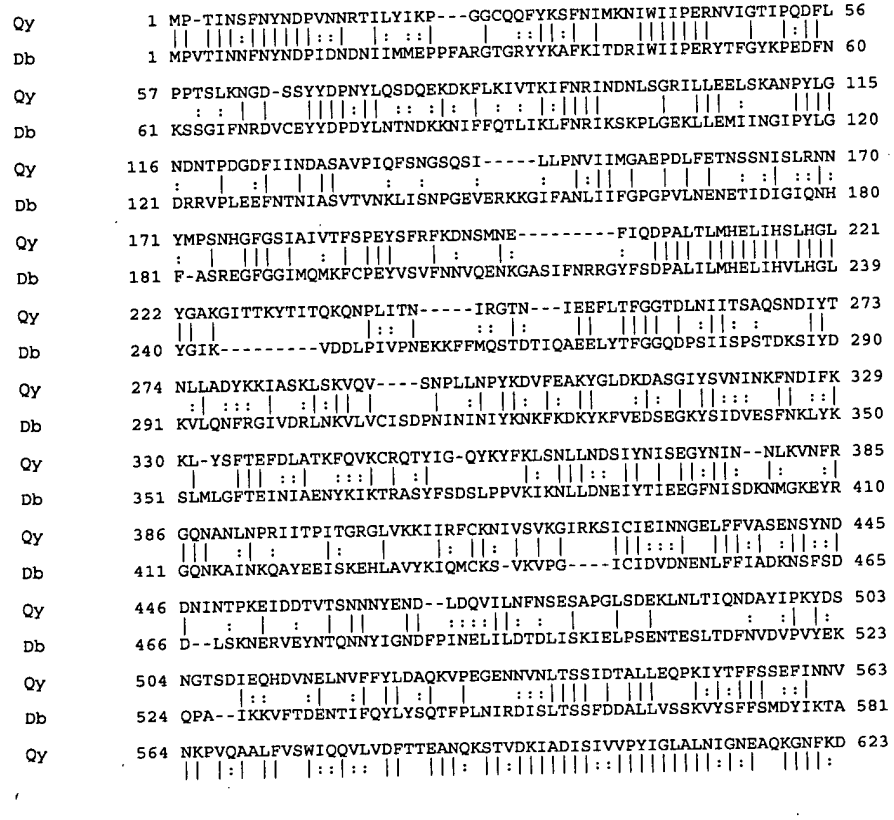
Qy       1199 FKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQ 1248
                |       ||      |:  ||      | ||   : :| |    |  | ||  : ||
Db       1237 TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 1296

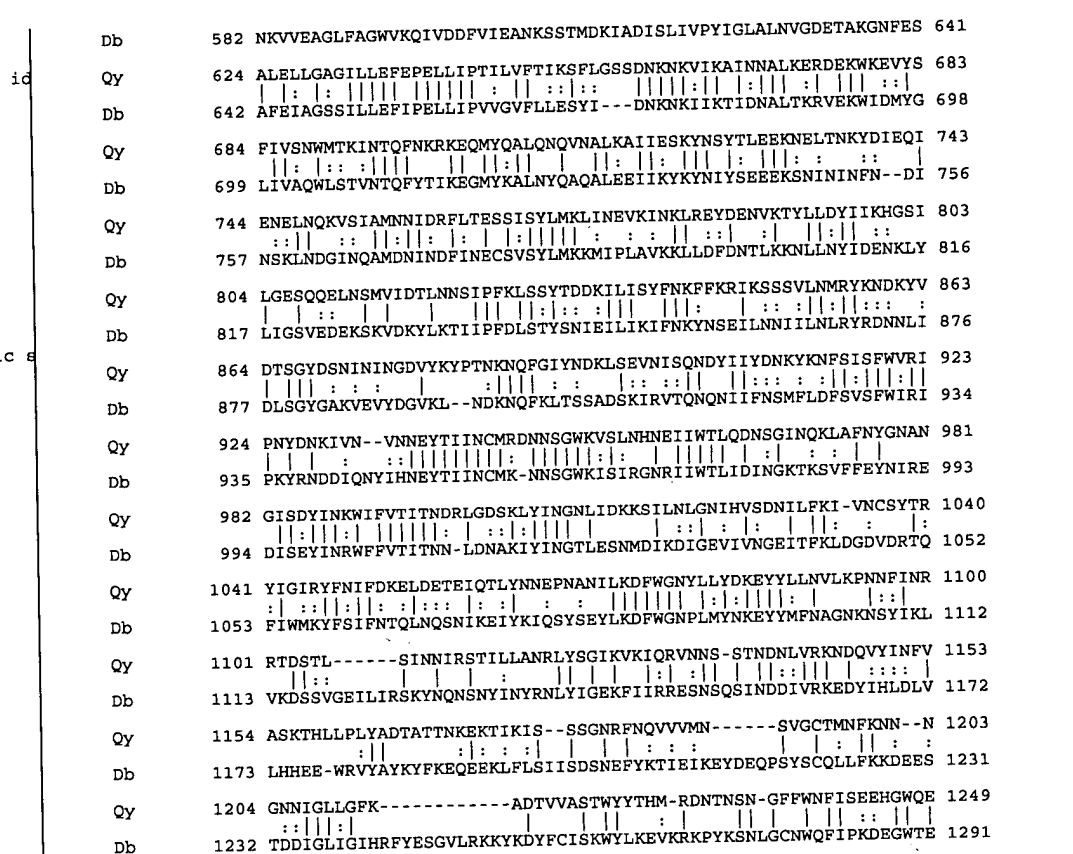
Qy       1249 E 1249
                |
Db       1297 E 1297


RESULT 8
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
```

R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 33.5%; Score 2181; DB 2; Length 1291;
Best Local Similarity 37.8%; Pred. No. 1.1e-81;
Matches 499; Conservative 263; Mismatches 458; Indels 100; Gaps 34;





RESULT 9
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publicat
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>

```
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match            33.4%;  Score 2176;  DB 1;  Length 1291;
  Best Local Similarity  37.4%;  Pred. No. 1.8e-81;
  Matches  491;  Conservative  267;  Mismatches  470;  Indels   84;  Gaps   32;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || |||:||||||::|  |: ::|    |  ::||:| |   |||||||   |  |:||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
                : : | |   ||||:|| :: :|: ||: : |:||||     |  ||| :    ||||

Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   | :|  | ||    :  :    :    |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   | :  |:          :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :    :  ::   : :      ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |     |  :| ||: |: ||   :|: | ||::: |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|        |: |||:: || | ||:||:  :::  :|||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:   |    |: ||   |||     |||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  ||| ::::||: :  |   |  |          :  :| |:      |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
               :|  :| || :|   |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::| |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              ||||| ||||||  :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751
              :||||   || ||:||   |  ||: ||: :|| |: :||: :    |   | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811
              : |::||: |:   |:|||| :  : : || ::|  :|  ||:|| ::   |  | :
Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
              | |   |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |     :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINK 989
               ::|||||||||: ||||||:|:  | ||||| | :|  : : | |      ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | ||||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS---- 1104
              ||: || :: |:  |  :  : ||||||| |:|:||||: |    |::|  : ||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLP 1161
                | |  |  |  :    || | |  |:| :|| | ||::||| | :|:::|  :
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWR 1179
```

```
Qy       1162 LYADTATTNKEKTI---KISSSGNRFNQVVVMN-----SVGCTMNFKNN--NGNNIGLLG 1211
                :|      :|: :    || |   :| : :       :  | : || :  : : |||:|
Db       1180 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 1239

Qy       1212 FK-----------ADTVVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249
                           |    | ||   ::    :   |  | || :: || |
Db       1240 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 10
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
C;Comment: dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized,
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of syr
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptc
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match   32.6%; Score 2120.5; DB 1; Length 1315;
  Best Local Similarity   35.6%; Pred. No. 3.4e-79;
  Matches 486; Conservative 262; Mismatches 444; Indels 175; Gaps   38;

```
Qy          1 MP-TINSFNYNDPVNNRTILYIKPGGCQQ---FYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || .|||:| |:|||| ||: ::|  |:    :||:| |   |||:|||   || |:||
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              ||:||  | | |||||||:|| :||:||: : |:|||| :|::|  ||:::  | |||||
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        117 DNTPDGDFIINDASAVPIQFS------NG--SQSILLPNVIIMGAEPDLFETNSSNISLR 168
                :    |   | ::  : |:      :|. ::| :| |:|| |  | | :     | ||
Db        121 SYSLLDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLR 177

Qy        169 ---NNYMPSNHGFGSIAIVTFSPEYSFRFKDN----------SMNEFIQDPALTLMHELI 215
                 || |   |||||  : | |||   | ||            ::: ||||| ||||||
Db        178 VDNKNYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELI 236

Qy        216 HSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275
              | |||||| : :::   | ||  : :    : ||  |||| | |:|:    ||:|
Db        237 HVLHGLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKT 295

Qy        276 LADYKKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-L 331
              | ||| ||:|||:|   |     :: || ::: ||  |||::| | || :||  ::   :
Db        296 LNDYKAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIM 355

Qy        332 YSFTEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388
              | ||| :|  || :| | :|    :   |: |||:|:||| :||:||  :||  ::|||
Db        356 YGFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQN 415

Qy        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGEL 434
                :|      : | ||| |:|  || |:     ||::                :||:| | :|
Db        416 MRVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDL 475

Qy        435 FFVASENSYNDDNINTPKEIDDTVTSNN-----NYENDLDQVILNFNSESAPGLSDEKLN 489
              |:| :||:::::      |: |: |       |:   ||::|:::| :|    | :::
Db        476 TFIAEKNSFSEEPFQ-----DEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTT 530

Qy        490 LTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQP 549
                :   | |:| ||  | || |::::  :: || ||| |     : :|:|:| ||:
Db        531 PVTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINST 590

Qy        550 KIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLA 609
              |||::| |  |: ||:  |  ||: |:: :: ||| |::||:|:|||:|:| :|||||
Db        591 KIYSYFPS-VISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPA 649

Qy        610 LNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINN 669
              |||  :  :|||  |||  |  :|||: ||: :| |   :|     ||  | |:|| |:|
Db        650 LNIVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAE---SSTQKEKIIKTIDN 706

Qy        670 ALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLE 729
              |::| ||| |||  : : |:  :|||| ||  |||::|: ||:|:| ||: :|  |:
Db        707 FLEKRYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGP 766

Qy        730 EKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLM-KLINEVKINKLREYDEN 788
              :| :: :  :|  ::|:| :| : || ||: |: ||| |:|: ::||| |  :| |:|
Db        767 DKEQIAD--EINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAK-KQLLEFDTQ 823
```

```
Qy        789 VKTYLLDYIIKHGSILGESQ-QELNSMVIDTLNNSIPFKLSSYTD---------DKILIS 838
Db        824 SKNILMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL-- 881

Qy        839 YFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSE 897
Db        882 ---------KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSE 932

Qy        898 VNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNV--NNEYTIINCMRDNN----SGWK 951
Db        933 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 992

Qy        952 VSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYI-NKWIFVTITNDRLGDSKLYINGN 1010
Db        993 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 1051

Qy       1011 LIDKKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069
Db       1052 LMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSI 1111

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLAN---------- 1119
Db       1112 TFLRDFWGNPLRYDTEYYLIPVAS--------SSKDVQLKNITDYMYLTNAPSYTNGKLN 1163

Qy       1120 ----RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLLPLYADTATTNKEK 1173
Db       1164 IYYRRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSY----------------NNNEH 1207

Qy       1174 TIKISSSGNRFNQV-----VVMNSVGCTMNFK-------------------NNNGNNIGL 1209
Db       1208 IVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGL 1267

Qy       1210 LGF---------KADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGW 1247
Db       1268 VGTHNGQIGNDPNRDILIASNWYFNHLKDKI--LGCDWYFVPTDEGW 1312
```

RESULT 11
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   28.7%; Score 1866; DB 2; Length 367;
  Best Local Similarity   97.8%; Pred. No. 1.5e-69;
  Matches 359; Conservative   5; Mismatches   3; Indels   0; Gaps   0;

```
Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELT 735
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLD 795
Db        121 NKYDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAF 975
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
Db        361 NYGNANG 367
```

RESULT 12
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   28.6%; Score 1863; DB 2; Length 1276;
  Best Local Similarity   33.7%; Pred. No. 1e-68;
  Matches 445; Conservative 259; Mismatches 492; Indels 126; Gaps   35;

```
Qy          4 INSFNYNDPVNNRTILYIK---------PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQD 54
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FLPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYL 114
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDG--DFIINDASAVPIQFSNGS---QSILLPNVIIMGAEPDLFETNSSNISLRN 169
Db        118 GDSSTPEDTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQ 177

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHG 220
Db        178 QSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQ 237

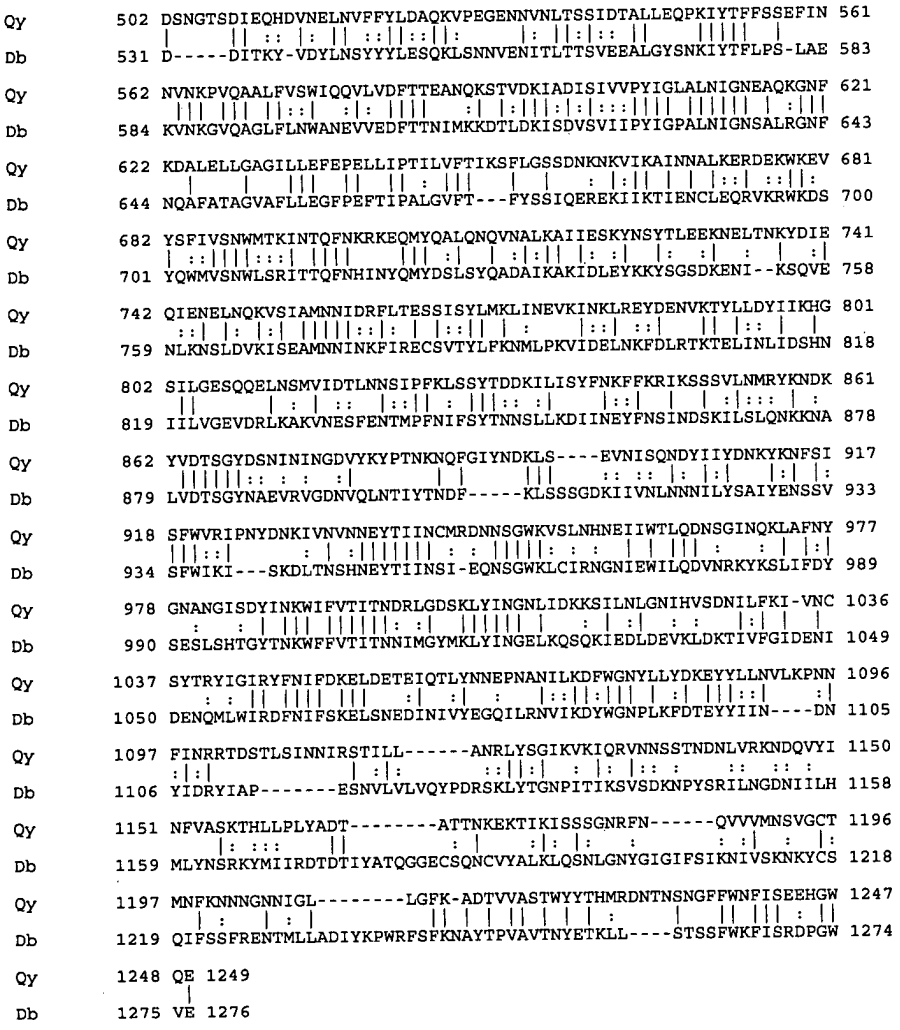
Qy        221 LYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
Db        238 LYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGH 296

Qy        279 YKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS 333
Db        297 YKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTN 356

Qy        334 -FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNAN 390
Db        357 VMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIE 416

Qy        391 LNPRIITPITGRGLVKKIIRFCKNIVSVKGIR-KSICIEINNGELFFVASENSYNDDNIN 449
Db        417 RNP-ALQKLSSESVVDLFTKVCLRL--TKNSRDDSTCIKVKNNRLPYVADKDSISQEIFE 473

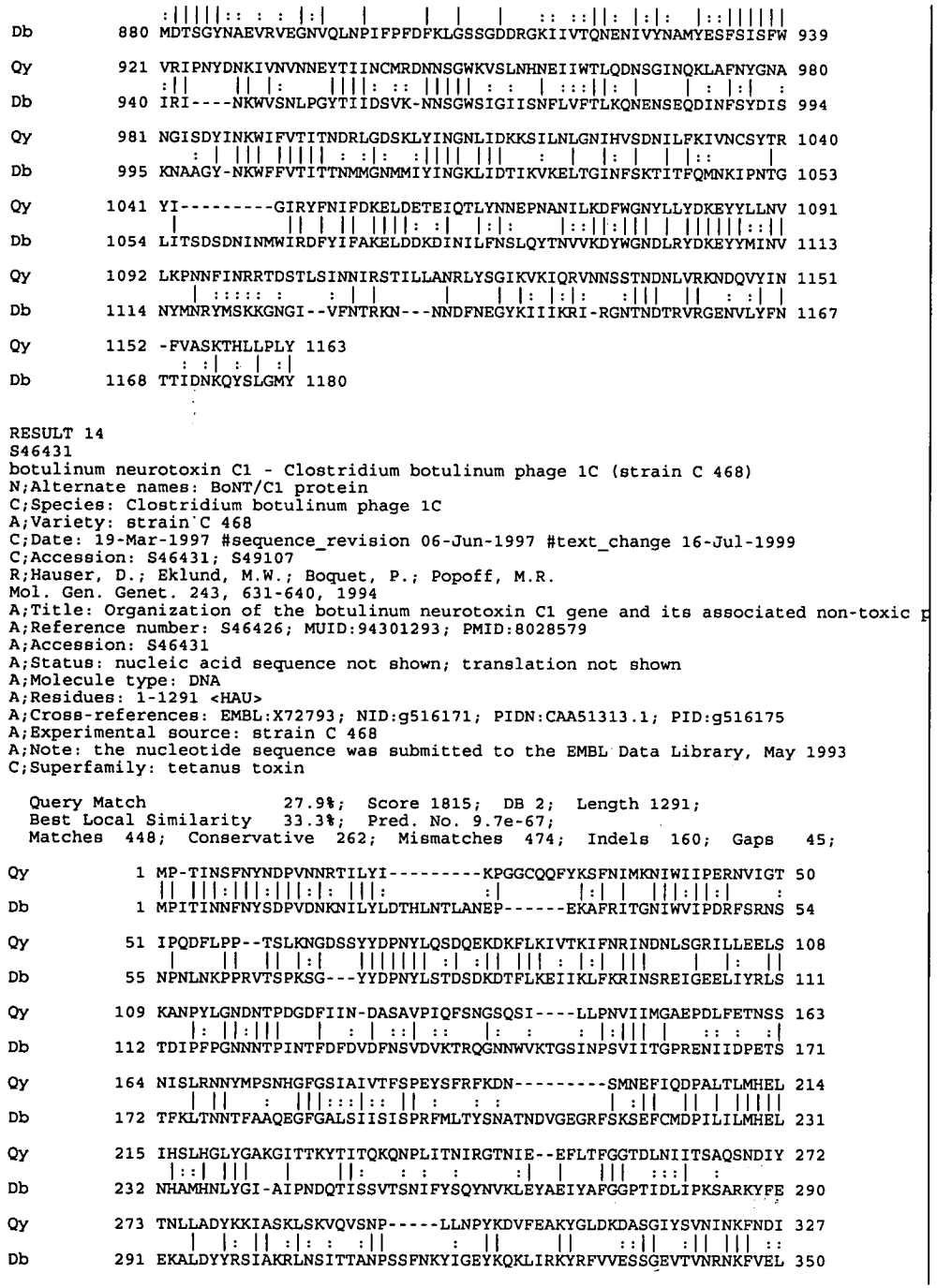
Qy        450 TPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPKY 501
Db        474 NKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVF---Y 530
```



RESULT 13
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>

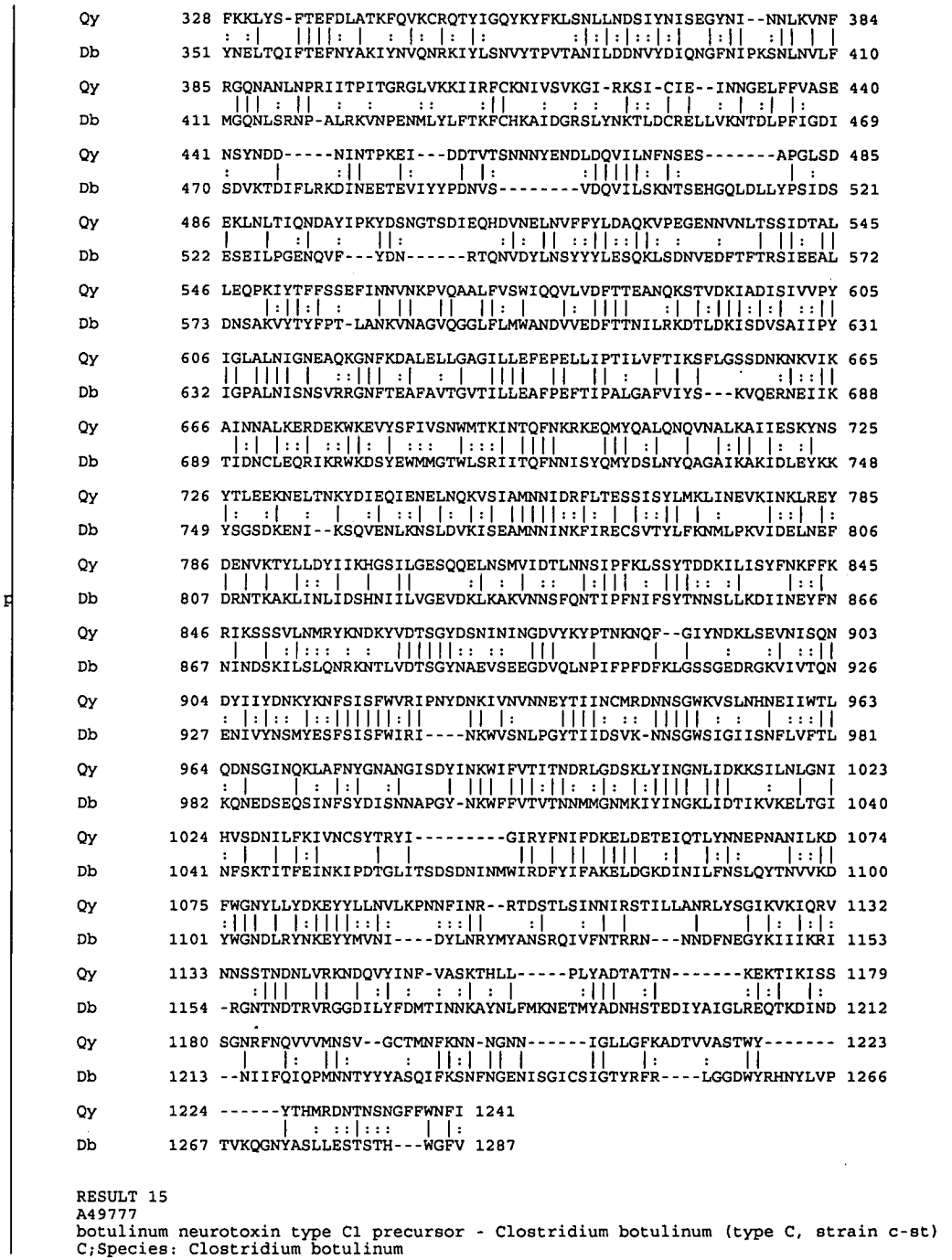
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membra
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

Query Match 28.4%; Score 1852; DB 2; Length 1285;
Best Local Similarity 35.3%; Pred. No. 3e-68;
Matches 428; Conservative 241; Mismatches 454; Indels 90; Gaps 30;



RESULT 14
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 27.9%; Score 1815; DB 2; Length 1291;
Best Local Similarity 33.3%; Pred. No. 9.7e-67;
Matches 448; Conservative 262; Mismatches 474; Indels 160; Gaps 45;



RESULT 15
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum

```
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            27.9%;  Score 1815;  DB 2;  Length 1291;
  Best Local Similarity  33.3%;  Pred. No. 9.7e-67;
  Matches  448;  Conservative  262;  Mismatches  474;  Indels  160;  Gaps   45;

Qy          1 MP-TINSFNYNDPVNNRTILYI---------KPGGCQQFYKSFNIMKNIWIIPERNVIGT 50
              || |||:|||:|||:|: |||:         :|       |:| |  |||:||:|    :
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEP------EKAFRITGNIWVIPDRFSRNS 54

Qy         51 IPQDFLPP--TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELS 108
               |    ||  || |:|   ||||||| :| :|| ||| : |:| |||     |  |:  ||
Db         55 NPNLNKPPRVTSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111

Qy        109 KANPYLGNDNTPDGDFIIN-DASAVPIQFSNGSQSI----LLPNVIIMGAEPDLFETNSS 163
              |: ||:|||   |  : | ::| ::   |:  :    : |:||| |   :: :  :|
Db        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171

Qy        164 NISLRNNYMPSNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHEL 214
                | ||   :  |||:::|:: || :   : :           | :|| || | |||||
Db        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231

Qy        215 IHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIY 272
              |::| |||    |    ||:   : :  :     :|  |    |||  :::|  :
Db        232 NHAMHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFE 290

Qy        273 TNLLADYKKIASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDI 327
                | |: || :|: :  :||      :  ||      ||    ::||  :|| ||| ::
Db        291 EKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL 350

Qy        328 FKKLYS-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNF 384
              : :|   ||||: |  : |: |: |:        :|:|:|::|:|  |:||  :|| | |
Db        351 YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLF 410

Qy        385 RGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASE 440
                ||| : ||  :  :    ::    :||   :  : :  |:: | |  : | :| |:
Db        411 MGQNLSRNP-ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDI 469

Qy        441 NSYNDD-----NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSD 485
              :    |     :||   |:    | |:         :|||||: |:           | :
Db        470 SDVKTDIFLRKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDS 521

Qy        486 EKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTAL 545
              |    |  :|  :   ||:          :|: || ::||::||: :    :   | ||: ||
Db        522 ESEILPGENQVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572

Qy        546 LEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPY 605
                  |:||:| :   | ||  ||  ||: |   |: ||||   :| |:|||:|:| ::||
Db        573 DNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPY 631

Qy        606 IGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIK 665
              || |||| |  ::||| :|  : |  ||||  ||  || :  | | |          :|::||
Db        632 IGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIK 688

Qy        666 AINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNS 725
              |:| |::| ::||: | :::  |:::| ||||    ||| :|  |  |:|| |: :|
Db        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748

Qy        726 YTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREY 785
              |:  :|  :  |  :| ::| |: |:| |||||::|: | |::|| | :    |::| |:
Db        749 YSGSDKENI--KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEF 806

Qy        786 DENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFK 845
              | | |  |:: |  |  ||     :| : | ::  |:||| : |||:: :|    |::|
Db        807 DRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFN 866

Qy        846 RIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQN 903
              |  | :|::: : :  ||||||:: ::  |||    |      |  |    :| ::||
Db        867 NINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQN 926

Qy        904 DYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTL 963
              : |:|:: |::||||||:||    || |:    ||||: :: ||||| : :  | :::||
Db        927 ENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTL 981

Qy        964 QDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNI 1023
              : |    | : |:|  :|    | ||| |||:||: :|: |:|||| |||    :  |  |
Db        982 KQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGI 1040

Qy       1024 HVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYNNEPNANILKD 1074
              : |  | |:|      |  |           || | || |||| :|  |:|:    |::||
Db       1041 NFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKD 1100

Qy       1075 FWGNYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRLYSGIKVKIQRV 1132
              :||| | |:|||::|:    :::||    :|  : | |        |  | |: |:|:
Db       1101 YWGNDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDFNEGYKIIIKRI 1153

Qy       1133 NNSSTNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTN-------KEKTIKISS 1179
                :|||  ||  | :| :  : :|: |      :|||  :|         :|:|  |:
Db       1154 -RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDIND 1212

Qy       1180 SGNRFNQVVVMNSV--GCTMNFKNN-NGNN------IGLLGFKADTVVASTWY------- 1223
                |   |:  ||:     :  ||:| || |      ||   |:   :  ||
Db       1213 --NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVP 1266

Qy       1224 ------YTHMRDNTNSNGFFWNFI 1241
                     |  : ::|:::   | |:
Db       1267 TVKQGNYASLLESTSTH---WGFV 1287


Search completed: September 10, 2004, 03:44:02
Job time : 27.0831 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 15.2517 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 6511
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6491.5	99.7	1250	1	BXE_CLOBU	P30995 clostridium
2	6308.5	96.9	1250	1	BXE_CLOBO	Q00496 clostridium
3	3986.5	61.2	1274	1	BXF_CLOBO	P30996 clostridium
4	2316	35.6	1295	1	BXA1_CLOBO	P10845 clostridium
5	2259	34.7	1295	1	BXA2_CLOBO	Q45894 clostridium
6	2230.5	34.3	1296	1	BXG_CLOBO	Q60393 clostridium
7	2174.5	33.4	1290	1	BXB_CLOBO	P10844 clostridium
8	2119	32.5	1314	1	TETX_CLOTE	P04958 clostridium
9	1863	28.6	1276	1	BXD_CLOBO	P19321 clostridium
10	1813.5	27.9	1290	1	BXC1_CLOBO	P18640 clostridium
11	765	11.7	1196	1	BXCN_CLOBO	P46081 clostridium
12	735	11.3	1162	1	BXEN_CLOBO	P46082 clostridium
13	727.5	11.2	1162	1	BXEN_CLOBU	Q06366 clostridium
14	269	4.1	2136	1	YCF2_MARPO	P09975 marchantia
15	266.5	4.1	1956	1	ATX1_PLAFA	Q04956 plasmodium
16	263.5	4.0	2710	1	TOXA_CLODI	P16154 clostridium
17	259.5	4.0	2867	1	RBP2_PLAVB	Q00799 plasmodium
18	256	3.9	2339	1	RPC1_PLAFA	P27625 plasmodium
19	250.5	3.8	1024	1	RPOB_PLAFA	P21421 plasmodium
20	250	3.8	2894	1	YD96_METJA	Q58791 methanococc
21	241	3.7	964	1	DPOL_CBEPV	P30319 choristoneu
22	241	3.7	1738	1	YCF1_EPIVI	Q00383 epifagus vi
23	240.5	3.7	2366	1	TOXB_CLODI	P18177 clostridium
24	238	3.7	2869	1	RBP1_PLAVB	Q00798 plasmodium
25	237	3.6	2971	1	YCX9_CHLRE	Q32065 chlamydomon
26	236	3.6	968	1	Y080_BUCAP	Q8ka32 buchnera ap
27	234.5	3.6	1061	1	EX5C_BUCAP	Q8k9b0 buchnera ap
28	232.5	3.6	1167	1	CAGA_HELPJ	Q9zlt1 helicobacte
29	232.5	3.6	1682	1	MSP1_PLAF3	P19598 plasmodium
30	230	3.5	760	1	ETF2_MSEPV	Q9yw29 melanoplus
31	230	3.5	1169	1	EX5B_BORBU	O51578 borrelia bu
32	230	3.5	1726	1	MSP1_PLAFP	P50495 plasmodium
33	228.5	3.5	1701	1	MSP1_PLAFF	P13819 plasmodium
34	228.5	3.5	2334	1	WAPA_BACSU	Q07833 bacillus su
35	228	3.5	1726	1	MSP1_PLAFC	P04934 plasmodium
36	226	3.5	1630	1	MSP1_PLAFK	P04932 plasmodium
37	225.5	3.5	1701	1	MSP1_PLAFM	P08569 plasmodium
38	225	3.5	1639	1	MSP1_PLAFW	P04933 plasmodium
39	224.5	3.4	971	1	Y228_BORBU	O51246 borrelia bu
40	224	3.4	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
41	221	3.4	2748	1	NUM1_YEAST	Q00402 saccharomyc
42	220	3.4	1111	1	KIP1_YEAST	P28742 saccharomyc
43	220	3.4	1271	1	Y338_MYCGE	P47580 mycoplasma
44	219.5	3.4	1186	1	CAGA_HELPY	P55980 helicobacte
45	218	3.3	1179	1	EX5B_BUCAP	Q8k9a9 buchnera ap

ALIGNMENTS

RESULT 1
BXE_CLOBU
ID BXE_CLOBU STANDARD; PRT; 1250 AA.
AC P30995;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43181, and ATCC 43755;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinal neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA Yokosawa N., Yashiki T., Oguma K.;
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.
RC STRAIN=5262;
RA Gimenez J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             99.7%;  Score 6491.5;  DB 1;  Length 1250;
  Best Local Similarity   99.8%;  Pred. No. 1e-254;
  Matches 1248;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy          2 PTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 60

Qy         62 KNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTPD 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTPD 120

Qy        122 GDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 180

Qy        182 AIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 240

Qy        242 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 300

Qy        302 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 360

Qy        362 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 420

Qy        422 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 480

Qy        482 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 540

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISI 600

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 660

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 720

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 780

Qy        782 LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 840

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 900

Qy        902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 960

Qy        962 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1020

Qy       1022 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLL 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLL 1080

Qy       1082 YDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 YDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1140

Qy       1142 VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-CTMNFK 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1141 VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250


RESULT 2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             96.9%;  Score 6308.5;  DB 1;  Length 1250;
  Best Local Similarity   96.9%;  Pred. No. 2.4e-247;
  Matches 1211;  Conservative   16;  Mismatches   22;  Indels    1;  Gaps    1;

Qy          2 PTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 61
              | |||||||||||:|||||||||||:||||||||||||||||||||| |||| |||||
Db          1 PKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSL 60

Qy         62 KNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTPD 121
              ||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 KNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPD 120

Qy        122 GDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 181
                | | ||||| |:|||||| |||||||||||||||||||||||||||||| ||||
Db        121 NQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHRFGSI 180

Qy        182 AIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 241
              |||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        181 AIVTFSPEYSFRFNDNCMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 240

Qy        242 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 300

Qy        302 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 361
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        301 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKLS 360

Qy        362 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 420

Qy        422 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 480

Qy        482 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 540

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 600
```

```
Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 660

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:| ||||
Db        661 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIES 720

Qy        722 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              ||||||||||||||||:|||||||||||||||||||||||||||||:||||||||
Db        721 KYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINK 780

Qy        782 LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 841
              ||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||||
Db        781 LREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFN 840

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 900

Qy        902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 960

Qy        962 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1021
              | :|| ||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        961 TFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLG 1020

Qy       1022 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLL 1081
              |||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db       1021 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLL 1080

Qy       1082 YDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1141
              ||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||||||
Db       1081 YDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1140

Qy       1142 VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-CTMNFK 1200
              |||||||||||||||||| ||||||||||||||||||||||||||||||||| ||||||
Db       1141 VRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNFK 1200

Qy       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 NNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1250
```

```
RESULT 3
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN        1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN      437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL      227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   228    228       BY SIMILARITY.
FT   METAL      231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID   429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             61.2%;  Score 3986.5;  DB 1;  Length 1274;
```

Best Local Similarity 60.3%; Pred. No. 9.8e-154;
Matches 772; Conservative 224; Mismatches 246; Indels 39; Gaps 17;

```
Qy          1 MP-TINSFNYNDPVNNRTILYIK---PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              ||  |||||||||||: ||||::       :::||:| ||:|:|||||| ||| | ||
Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTNPSDFD 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              || ||||| |:|||||||  :| ||||::||    |:| ||| | :|::||:|:|  | |||||
Db         61 PPASLKNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYLGN 120

Qy        117 DNTPDGDFI-INDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRN------ 169
              |:||  :|  :   ::| |: |   :| :| |::::|| ||:||  |    :|
Db        121 DHTPIDEFSPVTRTTSVNIKLSTNVESSMLLNLLVLGAGPDIFE--SCCYPVRKLIDPDV 178

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFKD------NSMNEFIQDPALTLMHELIHSLHGLYG 223
              | |||:||||| |||||||| : | |      :|   ||.|||::| |||||:||||||
Db        179 VYDPSNYGFGSINIVTFSPEYEYTFNDISGGHNSSTESFIADPAISLAHELIHALHGLYG 238

Qy        224 AKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIA 283
              |:|:| : ||  || ||:   :  :|||||||| ||||||||    || ||||:|:|||
Db        239 ARGVTYEETIEVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIA 298

Qy        284 SKLSKVQVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLAT 341
              ::||:|  : |   :| ||| |: ||||||:| | |:|| |||||:|:|||||||| |||
Db        299 TRLSEVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTESDLAN 358

Qy        342 KFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITG 401
              ||:|||| ||  :|:: |: |||:| || :|||:|| || || |||:  |||:||  |
Db        359 KFKVKCRNTYFIKYEFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQSIKLNPKIIDSIPD 418

Qy        402 RGLVKKIIRFCKNIVSVKGIR--KSICIEINNGELFFVASENSYNDDNINTPKEIDDTVT 459
              :|||:||::|||::  || :    :|| :|| ||||||||:|||:::||||||||||
Db        419 KGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNSELFFVASESSYNENDINTPKEIDDTTN 478

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519
              |||| |:||:|||::||:: | :|:  ||  :|:::|:|:|||||||:||::|| : ||
Db        479 LNNNYRNNLDEVILDYNSQTIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYDVVDFNV 538

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 579
              |||| |||||||| |::|||||||||||: |   |||||||: :|||| ||||: || :|
Db        539 FFYLHAQKVPEGETNISLTSSIDTALLEESK-DIFFSSEFIDTINKPVNAALFIDWISKV 597

Qy        580 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPE 639
              : |||||| ||||||||||||::|||:||||||  ||:||||::| |||| |||||| ||
Db        598 IRDFTTEATQKSTVDKIADISLIVPYVGLALNIIIEAEKGNFEEAFELLGVGILLEFVPE 657

Qy        640 LLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNK 699
              | || ||||||||:: | :|||| ||||||:| ||: ||||:||:|||||:|:|||||||
Db        658 LTIPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRINTQFNK 717

Qy        700 RKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNID 759
              |||||||||||||:|:|  || |||:|| :||| | ::|:|  || |||:|||:|| ||:
Db        718 RKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESEYNINNIEEELNKKVSLAMKNIE 777

Qy        760 RFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTL 819
              ||:|||||||||||||| |: ||::|| :||: ||:||: | |||||   ||: :|  ||
Db        778 RFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLLNYILDHRSILGEQTNELSDLVTSTL 837

Qy        820 NNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVY 879
              |:||||:|||||:||||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:||
Db        838 NSSIPFELSSYTNDKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVY 897

Qy        880 KYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTI 939
              | ||:|||||||| :|||||:||: |||:::|:|||||||||||  :  | :| | ||||
Db        898 IYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTI 956

Qy        940 INCMRDNNSGWKVSL---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTIT 996
              |||| :||||||:||      |||||||| ||  : | | |  | ||:|||||||||||
Db        957 INCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTIT 1016

Qy        997 NDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDE 1056
              |:|||:|::||||||| :||| |||:|||||||||||| |    |:||||| :|: |||:

Db       1017 NNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDK 1076

Qy       1057 TEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR---- 1112
              |||:|||:|||: :|||::|||||||:|:||| |:|: : :|   :   |:||  |
Db       1077 TEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS-GILNINQQRGVTE 1135

Qy       1113 STILLANRLYSGIKVKIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTN 1170
                :: |  :|| |::| |::    : |  || |||||  ||| | :     |||||  :
Db       1136 GSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSE 1193

Qy       1171 KEKTIKISSSGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMR 1228
              ||| |: |:  :   |::||:|:|  |||||:||||:||||||| :: :|||:||| ::|
Db       1194 KEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIR 1253

Qy       1229 DNTNSNGFFWNFISEEHGWQE 1249
                ||:||| ||: ||:|:||:|
Db       1254 RNTSSNGCFWSSISKENGWKE 1274
```

RESULT 4
BXA1_CLOBO
ID BXA1_CLOBO STANDARD; PRT; 1295 AA.
AC P10845; P01561; P18639;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 2916;
RX MEDLINE=90235864; PubMed=2185020;
RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type A
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 189:73-81(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=62A;
RX MEDLINE=90264400; PubMed=2160960;
RA Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [3]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=62A;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN=Hall;
RX MEDLINE=89350959; PubMed=2669749;
RA Betley M.J., Somers E., Dasgupta B.R.;
RT "Characterization of botulinum type A neurotoxin gene: delineation of
RT the N-terminal encoding region.";
RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN [5]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=Type A NIH;

```
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
```

```
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match           35.6%;  Score 2316;  DB 1;  Length 1295;
  Best Local Similarity  40.0%;  Pred. No. 2.1e-86;
  Matches  536;  Conservative  248;  Mismatches  419;  Indels  136;  Gaps   44;

Qy          2 PTIN-SFNYNDPVNNRTILYIK-PG-GCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DFL 56
              | :|  ||| ||||   | ||| |  |  |  |:| |   ||:||||:   | |:  |
Db          1 PFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLN 59

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              ||   |   ||||  || :| ||| :|| |||:| ||     ||:||  :  :  |: |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        117 DNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMPS 174
              :| | :  : | : : :   :||      |::|:|   |:  ||  |    : |    :
Db        120 -STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LT 175

Qy        175 NHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGAK 225
              :|:||   : |||:::| |:::            :|  |||:|| |||||: | ||
Db        176 RNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY--- 232

Qy        226 GITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADY 279
              ||        |   ||    : | :  ||  |||| |   | | | |:          :
Db        233 GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKF 289

Qy        280 KKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 335
              | ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|
Db        290 KDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT 349

Qy        336 EFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 392
              | :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|
Db        350 EDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEIN 408

Qy        393 PRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGEL 434
              | :   ||       :  | | :: |:||  |                 :||::|| :|
Db        409 NMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDL 461

Qy        435 FFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSD 485
              ||  ||::: :| :|  :||    ||: | |        : : |  | || :: |  :|
Db        462 FFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISI 516

Qy        486 EKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSID 542
              | |:  |       :|  :   ||    :::::::  :| || ||:   |:: : ||:|::
Db        517 ENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVN 572

Qy        543 TALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIV 602
              |||   ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|:
Db        573 EALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITII 632

Qy        603 VPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNK 662
              :|||| ||||||    | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVL 689

Qy        663 VIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESK 722
              ::  |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | ||||  :
Db        690 TVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ 749

Qy        723 YNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKL 782
              || || |||| :   ::|: : ::||: :: || ||::|| : |:|||  :    : :|
Db        750 YNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRL 807

Qy        783 REYDENVKTYLLDYII-KHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 841
              ::| ::|  || ||    |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 EDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFT 866

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
              :: | | ::|:||:||:::  :| | | | |||    |   | :|||  ::| : |:: :
Db        867 EYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVI 926

Qy        902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
              : |:|:: |:||| |||:||| | | | ::|||||||||| :|||||||||: ||||

Db        927 LKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIW 984

Qy        962 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1021
              ||||   | |:: | |    ||||||:||||||||:|| :||:|||| |||:| | |||
Db        985 TLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG 1044

Qy       1022 NIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||| |:||:||:   |  | ||| |:|||:|||||:| ||: ||:|: |: |||||||:||
Db       1045 NIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYL 1104

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNI---------------RSTILLANRLYSG 1124
              ||| ||:||:  || :::        :||:                 : | | : || |
Db       1105 QYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy       1125 IKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK-EKTIKISSSGN 1182
              |  |::   |   ||:|| ||:|||| |  :| : |   |  |   ::|  ||
Db       1157 TKFIIKKY-ASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVGN 1215

Qy       1183 RFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDNT 1231
              :||||| |          | || ::||||:|| :||         :||| ||   :  ::
Db       1216 -LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSS 1274

Qy       1232 NSNGFFWNFISEEHGWQEK 1250
                : |  | ||  : || |:
Db       1275 RTLGCSWEFIPVDDGWGER 1293


RESULT 5
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
```

```
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             34.7%;  Score 2259;  DB 1;  Length 1295;
  Best Local Similarity   38.7%;  Pred. No. 4.3e-84;
  Matches  522;  Conservative  240;  Mismatches  428;  Indels  160;  Gaps   38;

Qy          2 PTIN-SFNYNDPVNNRTILYIK--PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ--DFL 56
              | :|  ||| ||||   | |||    |  |  |:| |   ||:||||:   | |:  |
Db          1 PFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLN 59

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              ||   |    ||||  || :| ||| :|| |||:| ||     ||:||  : :  |: |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        117 DNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNNYMPS 174
               :| | :  : | : :  :||      |::|:|   |:  ||  |    : |    :
Db        120 -STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHDVLN---LT 175

Qy        175 NHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHSLHGLYGAK 225
               :|:||   : |||:::| |:::           :|  |||:|| |||||: | ||
Db        176 RNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAEHRLY--- 232

Qy        226 GITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADY 279
              ||         | |   ||    : |  :  ||  |||| |   | | | |:       :
Db        233 GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKF 289

Qy        280 KKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 335
               | :|| |:| :    :    |   |:||: || | :| || :||:  ||: ::| |   :|
Db        290 KDVASTLNKAKSIIGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT 349

Qy        336 EFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
              | :    |:|  |:||:   |     |:: |  | | :|:|:   ||  || |||  :|

Db        350 EDNFVNFFKVINRKTYLNFDKAVFRINIVPDENYTIKDGFNLKGANLSTNFNGQNTEINS 409

Qy        394 RIITPI---TGRGLVKKIIRFCKNIVSVKGI--------------RKSICIEINNGELF 435
              |  | :   ||      :  | | :: |:||                 :|||:|| :||
Db        410 RNFTRLKNFTG------LFEFYK-LLCVRGIIPFKTKSLDEGYNKALNDLCIKVNNWDLF 462

Qy        436 FVASENSYNDDNINTPKEID--DTVTSNNNYE--------NDLDQVILNFNSESAPGLSD 485
              |  ||::: :|      :|  : :|:: | |         : :  | | |: :: |
Db        463 FSPSEDNFTND-------LDKVEEITADTNIEAAEENISLDLIQQYYLTFDFDNEPE--- 512

Qy        486 EKLNLTIQNDAYIPKYDSNGTSDI----------------EQHDVNELNVFFYLDAQKVP 529
                |::|:          | :|||                :::::::  :| || ||:
Db        513 ---NISIE----------NLSSDIIGQLEPMPNIERFPNGKKYELDKYTMFHYLRAQEFE 559

Qy        530 EGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQ 589
               |:: : ||:| : |||:    ||||||:::  :|| |:| :|::| :::: ||| | |:
Db        560 HGDSRIILTNSAEEALLKPNVAYTFFSSKYVKKINKAVEAFMFLNWAEELVYDFTDETNE 619

Qy        590 KSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFT 649
               :|:|||||:|:|||| ||||||   || | :|:   |   :||| ||  :|   |
Db        620 VTTMDKIADITIIVPYIGPALNIGNMLSKGEFVEAIIFTGVVAMLEFIPEYALPVFGTFA 679

Qy        650 IKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQ 709
              | |::    ||   :: ||||| :|:||| ||| : |:||: |:||| :  :|:| :||:
Db        680 IVSYIA---NKVLTVQTINNALSKRNEKWDEVYKYTVTNWLAKVNTQIDLIREKMKKALE 736

Qy        710 NQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISY 769
              ||  | ||||  :|| || |||| :   ::|: : ::||: :: || ||::|| : |:||
Db        737 NQAEATKAIINYQYNQYTEEEKNNI--NFNIDDLSSKLNESINSAMININKFLDQCSVSY 794

Qy        770 LMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSS 829
              ||  :   : :|:::| :|:  || ||  :   |       |   | :||:  |||:||
Db        795 LMNSMIPYAVKRLKDFDASVRDVLLKYIYDNRGTLVLQVDRLKDEVNNTLSADIPFQLSK 854

Qy        830 YTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFG 889
              | |:| |:|  | :: | | ::|:|:: || |  :| | | : |||   ||    :|||
Db        855 YVDNKKLLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIK 914

Qy        890 IYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSG 949
              : | : | : :  : |:|:: |:||| |||::|| | :|| |:||||||||: :||||
Db        915 LINLESSTIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKI-NLNNEYTIINCI-ENNSG 972

Qy        950 WKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYING 1009
              |||||: |||||||    |:: | |   |||||:|||||||:||  ||:||||
Db        973 WKVSLNYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYING 1032

Qy       1010 NLIDKKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYNNEPN 1068
               |||:| | |||||| |: |:||:  |    ||| |:|||:|||||:| ||: ||::: |
Db       1033 RLIDQKPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSN 1092

Qy       1069 ANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNI----------------R 1112
              : |||||||||| ||| ||:||: || :::         :|||
Db       1093 SGILKDFWGNYLQYDKPYYMLNLFDPNKYVD--------VNNIGIRGYMYLKGPRGSVVT 1144

Qy       1113 STILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVA-SKTHLLPLYADTATTNK 1171
              : | | : || | |  |::   |  ||:|| ||:|||| |  :| : |   |  |   |
Db       1145 TNIYLNSTLYEGTKFIIKKY-ASGNEDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEK 1203

Qy       1172 -EKTIKISSSGNRFNQVVVMNSVG-------CTMNFKNNNGNNIGLLGF----KADTVVA 1219
                ::|   ||  :|||| |         | || ::||||:|| :||       :||
Db       1204 ILSALEIPDVGN-LSQVVVMKSKDDQGIRNKCKMNLQDNNGNDIGFIGFHLYDNIAKLVA 1262

Qy       1220 STWYYTHMRDNTNSNGFFWNFISEEHGWQE 1249
              | ||   :  : : |  | ||  : || |
Db       1263 SNWYNRQVGKASRTFGCSWEFIPVDDGWGE 1292


RESULT 6
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
```

```
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             34.3%;  Score 2230.5;  DB 1;  Length 1296;
  Best Local Similarity   38.6%;  Pred. No. 6e-83;
  Matches  509;  Conservative  247;  Mismatches  466;  Indels   95;  Gaps   33;

Qy          4 INSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              |  ||||||:||  |: ::|    |   :||:| |:  |||:|||   |  |  |   |
Db          4 IKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 63

Qy         61 LKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNT 119
              : : |   |||| ||::| ||||||| : |:|||||   ||: ||: :  | ||||| :|
Db         64 VFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAST 123

Qy        120 PDGDFIINDASAVPIQ---FSNGSQSI---LLPNVIIMGAEPDLFETNSSNISLRNNYMP 173
              |   |  | |: | |       |::     |: |:|| |  | | : | ::  : | : |
Db        124 PPDKFAANVAN-VSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNGHSP 181

Qy        174 SNHGFGSIAIVTFSPEYSFRF------KDNSM---NEFIQDPALTLMHELIHSLHGLYGA 224
              : |||:  :: | |      |      || |:     :  ||||||||||| ||||||
Db        182 ISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGI 241

Qy        225 K----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
              |    ||  ||    | :|:         ||  |||| | ::|: :   :||   | :
Db        242 KISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKALQN 294

Qy        279 YKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSF 334
              :: ||::|   |  | |   :: || ::: ||   :| :| |||: :||: ::| | : |
Db        295 FQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGF 354

Qy        335 TEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :|| :: :| | :|  :|     |  ||:::||  :||:||   |||  | |||  :
Db        355 TETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAV 414

Qy        392 NPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTP 451
              |      |:  ||   | || ::     :  || :|| :|||:|:::|:: |
Db        415 NKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKA- 473

Qy        452 KEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYDSNG 505
                  :|:   | ||  ||:  :||:||: :  |   | :|     |  :|  || |
Db        474 ----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQ- 528

Qy        506 TSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNK 565
              | :::  |:  ::| || ||  |    |: ||:|:: ||    |:|||||:  :   |
Db        529 -SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANT 587

Qy        566 PVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDAL 625
              | |:|||:|:: |: |||:|: ||||:||::|:|::||||  |||:|||  | |||:|
Db        588 VVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAF 647

Qy        626 ELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI 685
              |: || ||:|| |||::| :  ||::|::|   ||  :|  |:||||:||:|| ::|  |
Db        648 EIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGLI 704

Qy        686 VSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIEN 745
              || |:: :||||   ||:|| || ||  |:: ||| :|| |: |:|  :    |   |:
Db        705 VSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFNDIDF 762

Qy        746 ELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILG 805
              :||| :::|:|||| |: : |||||   :  : : ||:::|:|:|  ||:||  :   |
Db        763 KLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLL 822

Qy        806 ESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDT 865
              :    | | |   | :|||| || || | |||  || :   | |:::|:: |:  : :|:
Db        823 DEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDS 882

Qy        866 SGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPN 925
              ||| : :|:  ||         || : | : | :   |: :::||: : ||||:|||| |
Db        883 SGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPK 942

Qy        926 YDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGI 983
              |:|  :   : |||||:|:: |:|||||:  | ||||| |  :  :: : | |   : |
Db        943 YNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNI 1001

Qy        984 SDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCS-YTRYI 1042
              ||||||  :|||||||:: :||||:|   : ||||  |: |::| ||::||:  |:::
Db       1002 SDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFV 1061

Qy       1043 GIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFI---- 1098
              |: |||| :||: ||: :||  : : | ||||||  | || :||| |    | :|
Db       1062 WIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS 1121

Qy       1099 NRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN----- 1151
```

```
                |    |  :  :    || |::  |:: :||    |||:||: | :|:|
Db       1122 KASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNIS 1181

Qy       1152 --------FVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMN-SVGCTMNFKNN 1202
                     | ||     |:   |  | : |         :: :    |  : |  : |
Db       1182 DESYRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTF 1239

Qy       1203 NGNNIGLL----GFKADT----VVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249
                  ||      |:  ||      | ||   : :| |    |  | ||  : || |
Db       1240 GLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1296
```

```
RESULT 7
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             33.4%;  Score 2174.5;  DB 1;  Length 1290;
  Best Local Similarity   37.4%;  Pred. No. 1.1e-80;
```

```
  Matches  489;  Conservative  267;  Mismatches  470;  Indels   83;  Gaps   31;

Qy          3 TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPT 59
              |||:||||||::|  |: ::|    |  ::||:| |   ||||||    |  |:||   :
Db          3 TINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNKSS 62

Qy         60 SLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDN 118
               : | |   ||||:|| :: :|: ||: : |:||||     |  ||| :    ||||:
Db         63 GIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLGDRR 122

Qy        119 TPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNNYMP 173
              |  :|  | ||    :  :    :     :  |:|| |  | | |  : :| ::|::
Db        123 VPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHF-A 181

Qy        174 SNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGLYGA 224
              |  ||| |  : | |||   | :   |:         :  |||| ||||||| ||||||
Db        182 SREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLYGI 241

Qy        225 KGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIAS 284
              | :      :  ::   : :       ||  |||| | :||| :    ||  :| ::: |
Db        242 K-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIVD 300

Qy        285 KLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTEFDL 339
              :|:|| |      |  :| ||: |: ||   :|: | ||:::  |: ::| | : ||| ::
Db        301 RLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETNI 360

Qy        340 ATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNPRII 396
              |  :::| | :|        |: |||:: || | ||:||:  :::  :|||  :| :
Db        361 AENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQAY 420

Qy        397 TPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDD 456
              |:   |    |: ||   |||     |||:::| :|||:| :||::|| ::  : |:
Db        421 EEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNERIEY 474

Qy        457 TVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV 514
              | :|  |||  ::::||: :  |   |  |         :  :| |:      |::
Db        475 N-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKKIFT 531

Qy        515 NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVS 574
              :|  :| || :|  |    :::|||| | |||    |:|:||| ::|   || |:| ||
Db        532 DENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAG 591

Qy        575 WIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILL 634
              |::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::|  |: || |||
Db        592 WVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILL 651

Qy        635 EFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKIN 694
              || |||||| :  | ::|::   ||||:|| |:||| :|:||| ::|  ||: |:: :|
Db        652 EFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVN 708

Qy        695 TQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIA 754
              |||   || ||:||  |  ||: ||: :|| |: :||: :   |   | ::||: :: |
Db        709 TQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGINQA 766

Qy        755 MNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSM 814
              ::||: |:   |:||||| :  : : || ::|  :|  ||:|| ::   |  | :   |
Db        767 IDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSK 826

Qy        815 VIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINI 874
              |  |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| : : :
Db        827 VNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEV 886

Qy        875 NGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN-- 932
              |    :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :
Db        887 YDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNY 944

Qy        933 VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIF 992
              ::|||||||||: |||||:|:  | ||||| | :|  : : | |     ||:|||:| |
Db        945 IHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFF 1003

Qy        993 VTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFD 1051
              |||||: | ::|:|||| |     | ::  :  :  |:||:  :   |::| ::||:||:
Db       1004 VTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFN 1062

Qy       1052 KELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS------T 1105
              || :: |:  |  :  :  |||||| |:|:|||| |    |::|  : ||      |
Db       1063 TELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILT 1122

Qy       1106 LSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLPLYA 1164
              |  |  |  :   || | |  |:| :|| | ||::||| | :|::|  :      :|
Db       1123 RSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYT 1181

Qy       1165 DTATTNKEKTI---KISSSGNRFNQVVVMN-----SVGCTMNFKNN--NGNNIGLLGFK- 1213
                    :|: :    || |   :| : :       :  | : || :  : : |||:|
Db       1182 YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHR 1241

Qy       1214 ----------ADTVVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249
                        |   | ||  ::   :   |  | || :: || |
Db       1242 FYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1290


RESULT 8
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
```

```
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
```

```
FT   STRAND     1190   1190

  Query Match             32.5%;  Score 2119;  DB 1;  Length 1314;
  Best Local Similarity   35.5%;  Pred. No. 1.9e-78;
  Matches  484;  Conservative  262;  Mismatches  444;  Indels  174;  Gaps   37;

Qy          3 TINSFNYNDPVNNRTILYIKPGGCQQ---FYKSFNIMKNIWIIPERNVIGTIPQDFLPPT 59
              |||:| |:|||| ||: ::|  |:    :||:| |   |||:|||   || |:|| ||:
Db          3 TINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFNPPS 62

Qy         60 SLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNT 119
              ||  | | |||||||::| :||:||: : |:|||| :|::|  ||:::  | |||||  :
Db         63 SLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGNSYS 122

Qy        120 PDGDFIINDASAVPIQFS------NG--SQSILLPNVIIMGAEPDLFETNSSNISLR--- 168
                  |   | ::  : |:      :|  ::| :| |:|| |  | | :       | ||
Db        123 LLDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDN 179

Qy        169 NNYMPSNHGFGSIAIVTFSPEYSFRFKDN----------SMNEFIQDPALTLMHELIHSL 218
              || |   |||||  : | |||   | ||           ::: ||||| ||||||| |
Db        180 KNYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELIHVL 238

Qy        219 HGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
              ||||| : :::   |  ||   : :    : ||  |||| | |:|:    ||:|   | |
Db        239 HGLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLND 297

Qy        279 YKKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-LYSF 334
              || ||:|||:|   |     :: || ::: ||  |||::| | || :|| ::   :| |
Db        298 YKAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGF 357

Qy        335 TEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :|  || :| | :|     :   |: |||:|:||| :||:||   :||  ::|||  :
Db        358 TEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRV 417

Qy        392 NPRIITPITGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGELFFV 437
              |      : | ||| |:|  || |:    ||::               :||:| | :| |:
Db        418 NTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFI 477

Qy        438 ASENSYNDDNINTPKEIDDTVTSNN-----NYENDLDQVILNFNSESAPGLSDEKLNLTI 492
              | :||:::::       |: |: |      |:   ||::|:::| :|   | :::
Db        478 AEKNSFSEEPFQ-----DEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVT 532

Qy        493 QNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIY 552
              :  | |:| ||  | || |::::  :: || ||| |     : :|:|:| ||:   |||
Db        533 KGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIY 592

Qy        553 TFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNI 612
              ::| |  |: ||:  |  ||: |:: :: ||| |::||:|:|||:|:| :||||| ||||
Db        593 SYFPS-VISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNI 651

Qy        613 GNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALK 672
                :  :||| ||| |  :|||: ||: :| |  :|     ||  | |:|| |:| |:
Db        652 VKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAE---SSTQKEKIIKTIDNFLE 708

Qy        673 ERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKN 732
              :| ||| ||| : : |:  :|||| ||  |||::|: ||:|:| || :|  |:  :|
Db        709 KRYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKE 768

Qy        733 ELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLM-KLINEVKINKLREYDENVKT 791
              :: :  :|  ::|:| :| : || ||: |: ||| |:|: ::||| |  :| |:|   |
Db        769 QIAD--EINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAK-KQLLEFDTQSKN 825

Qy        792 YLLDYIIKHGSILGESQ-QELNSMVIDTLNNSIPFKLSSYTD---------DKILISYFN 841
              |: ||  :   :| :: ::| | :    :  |||  |   |          | ||
Db        826 ILMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL----- 880

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSEVNI 900
                  | |::||:   ||    | ||::|::     |    |  |     : |:: ||| :
Db        881 ------KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIV 934

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNV--NNEYTIINCMRDNN----SGWKVSL 954
              :   | |:: : ||::|||:|:|      :     |||:|| |: ::    ||| |||
Db        935 HKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSL 994

Qy        955 NHNEIIWTLQDNSGINQKLAFNYGNANGISDYI-NKWIFVTITNDRLGDSKLYINGNLID 1013
                | :||||:|::|  ::: |     :  : |: |||:|:|||||||  : ||||| |:
Db        995 KGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMG 1053

Qy       1014 KKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANIL 1072
                |  || |   :||  |:  | :  :|: |  | || | |:  ||: || :  :   |
Db       1054 SAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFL 1113

Qy       1073 KDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLAN------------- 1119
                :||||| | || ||||: |          :   : : ||   : | |
Db       1114 RDFWGNPLRYDTEYYLIPVAS--------SSKDVQLKNITDYMYLTNAPSYTNGKLNIYY 1165

Qy       1120 -RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLLPLYADTATTNKEKTIK 1176
                |||:|:|  |:|   ::  |: |:  |  ::|:::                 | |  :
Db       1166 RRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSY----------------NNNEHIVG 1209

Qy       1177 ISSSGNRFNQV-----VVMNSVGCTMNFK------------------NNNGNNIGLLGF 1212
                 || || :     |  |: |  :  |                   ::   ::||:|
Db       1210 YPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGT 1269

Qy       1213 ---------KADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGW 1247
                      | ::|| ||: |::|     |  | |:  : ||
Db       1270 HNGQIGNDPNRDILIASNWYFNHLKDKI--LGCDWYFVPTDEGW 1311


RESULT 9
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
```

```
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match           28.6%;  Score 1863;  DB 1;  Length 1276;
  Best Local Similarity  33.7%;  Pred. No. 3.8e-68;
  Matches  445;  Conservative  259;  Mismatches  492;  Indels  126;  Gaps   35;

Qy          4 INSFNYNDPVNNRTILYIK---------PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQD 54
              :  |||:||||:  |||::         |       |:| | :|||:||||    | |
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FLPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYL 114
                ||       |||||:|| :|::|| ||| : |:| |||:   |: |:  |   :|::
```

```
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDG--DFIINDASAVPIQFSNGS---QSILLPNVIIMGAEPDLFETNSSNISLRN 169
              |: :||:   ||  :  :    :| |||    :|: |:|:|  |  |:: :  :|
Db        118 GDSSTPEDTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQ 177

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHG 220
                ||  |||:::|:  :||:   | | : |:          |  || : ||||| ||||
Db        178 QSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQ 237

Qy        221 LYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
              |||   |  :   :   : | |:  ||  |||| |: ||   : : :     |
Db        238 LYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGH 296

Qy        279 YKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS 333
              || || :|:  :  : |      :: || :|  ||  ||| :| : |||:||| ::  | :
Db        297 YKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTN 356

Qy        334 -FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNAN 390
               :|   :::: || | |  ::     :|:|:|:|| | :|:|: |   |    |||
Db        357 VMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIE 416

Qy        391 LNPRIITPITGRGLVKKIIRFCKNIVSVKGIR-KSICIEINNGELFFVASENSYNDDNIN 449
              ||  :  ::   :|    : | :   |  | | ||:: |  | :|| ::|  : :
Db        417 RNP-ALQKLSSESVVDLFTKVCLRL--TKNSRDDSTCIKVKNNRLPYVADKDSISQEIFE 473

Qy        450 TPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPKY 501
                 |:|   | : :  ||: ||:      | |       | :: | ||| :   :  |
Db        474 NKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVF---Y 530

Qy        502 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 561
              |     || :: |: || ::||::||:      |: ||:|:: ||     |||||  |
Db        531 D-----DITKY-VDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPS-LAE 583

Qy        562 NVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 621
              ||| ||| ||::|  :|: ||||   :| |:|||:|:|:::|||| |||||| | :|||
Db        584 KVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNF 643

Qy        622 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 681
              |     |  ||| ||  || : |||   |  |   : |:|| | | |::| ::||:
Db        644 NQAFATAGVAFLLEGFPEFTIPALGVFT---FYSSIQEREKIIKTIENCLEQRVKRWKDS 700

Qy        682 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIE 741
              | ::||||:::| ||||     ||| :|  | :|:|| |: :|  |:  :|  :  |  :|
Db        701 YQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENI--KSQVE 758

Qy        742 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHG 801
              ::| |: |:| |||||::|: | |::|| | :     |::| ::|   || |::  |  |
Db        759 NLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHN 818

Qy        802 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 861
              ||      | : | ::  |::|| : |||:: :|    |::|  |  | :|::  | :
Db        819 IILVGEVDRLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNA 878

Qy        862 YVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS----EVNISQNDYIIYDNKYKNFSI 917
              ||||||:: : :  :|       | |      |||    :: :: |: |:|   |:| |:
Db        879 LVDTSGYNAEVRVGDNVQLNTIYTNDF-----KLSSSGDKIIVNLNNNILYSAIYENSSV 933

Qy        918 SFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNY 977
              |||::|     : | :|||||| : : |||||: : :  | | ||| :   : | |:|
Db        934 SFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDY 989

Qy        978 GNANGISDYINKWIPVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNC 1036
              :   : | ||| ||||||: :|  |||||| |   : | :|  : :    |:| |  |-
Db        990 SESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENI 1049

Qy       1037 SYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNN 1096
                : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|    :|
Db       1050 DENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DN 1105

Qy       1097 FINRRTDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYI 1150
              :|:|              | :|:      ::||:|  :  |: |:: :   ::  :: : :
Db       1106 YIDRYIAP-------ESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILH 1158
```

```
Qy       1151 NFVASKTHLLPLYADT--------ATTNKEKTIKISSSGNRFN------QVVVMNSVGCT 1196
                |: :::    ||         : |    :|: |:   :       : :|  :  |:
Db       1159 MLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCS 1218

Qy       1197 MNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGW 1247
                | :   | : |          || | | || | | | :     |   || ||| : ||
Db       1219 QIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFISRDPGW 1274

Qy       1248 QE 1249
                |
Db       1275 VE 1276


RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             27.9%;  Score 1813.5;  DB 1;  Length 1290;
  Best Local Similarity   33.3%;  Pred. No. 3.8e-66;
  Matches  446;  Conservative  262;  Mismatches  474;  Indels  159;  Gaps   44;

Qy         3 TINSFNYNDPVNNRTILYI---------KPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ 53
             |||:|||:|||:|: |||:         :|       |:| |  |||:||:|    : |
Db         3 TINNFNYSDPVDNKNILYLDTHLNTLANEP------EKAFRITGNIWVIPDRFSRNSNPN 56

Qy        54 DFLPP--TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKAN 111
               ||  || |:|   |||||||:| :|| ||| : |:| |||     |  |:  ||
Db        57 LNKPPRVTSPKSG---YYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDI 113

Qy       112 PYLGNDNTPDGDFIIN-DASAVPIQFSNGSQSI----LLPNVIIMGAEPDLFETNSSNIS 166
             |: ||:|||   |  : | ::| ::   |:  :    : |:||| |   :: :  :|
Db       114 PFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFK 173

Qy       167 LRNNYMPSNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHS 217
             | ||   :  |||:::|:: || :   : :           | :||  || | ||||| |:
Db       174 LTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHA 233

Qy       218 LHGLYGAKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNL 275
             :| |||   |    ||:   : :  :     :|  |   |||  :::|  :
Db       234 MHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKA 292

Qy       276 LADYKKIASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK 330
             |  |: || :|: :  :||      :  ||      ||    ::||  :|| ||| ::: :
Db       293 LDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNE 352

Qy       331 LYS-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQ 387
             |   ||||: |  : |: |: |:        :|:|:|::|:|  |:||  :|| | | ||
```

```
Db        353 LTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQ 412

Qy        388 NANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASENSY 443

Db        413 NLSRNP-ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDV 471

Qy        444 NDD-----NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSDEKL 488

Db        472 KTDIFLRKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDSESE 523

Qy        489 NLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548

Db        524 ILPGENQVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNS 574

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608

Db        575 AKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGP 633

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668

Db        634 ALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIKTID 690

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTL 728

Db        691 NCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSG 750

Qy        729 EEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788

Db        751 SDKENI--KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRN 808

Qy        789 VKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIK 848

Db        809 TKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNIN 868

Qy        849 SSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYI 906

Db        869 DSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENI 928

Qy        907 IYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDN 966

Db        929 VYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQN 983

Qy        967 SGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVS 1026

Db        984 EDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFS 1042

Qy       1027 DNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYNNEPNANILKDFWG 1077

Db       1043 KTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWG 1102

Qy       1078 NYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNS 1135

Db       1103 NDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RG 1154

Qy       1136 STNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTN-------KEKTIKISSSGN 1182

Db       1155 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDIND--N 1212

Qy       1183 RFNQVVVMNSV--GCTMNFKNN-NGNN------IGLLGFKADTVVASTWY---------- 1223

Db       1213 IIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVK 1268

Qy       1224 ---YTHMRDNTNSNGFFWNFI 1241

Db       1269 QGNYASLLESTSTH---WGFV 1286


RESULT 11
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match           11.7%;  Score 765;  DB 1;  Length 1196;
  Best Local Similarity  24.4%;  Pred. No. 6.4e-24;
  Matches  320;  Conservative  224;  Mismatches  458;  Indels  310;  Gaps   64;

Qy          5 NSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSLKNG 64

Db          5 DDLNINSPVDNKNVVIVRARKTNTFFKAFKVAPNIWVAPER-YYGE-PLDIAEEYKL--- 59

Qy         65 DSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANP----YLGN---- 116

Db         60 DGGIYDSNFLSQDSERENFLQAIIILLKRINNTISGKQLLSLISTAIPFPYGYIGGGYSS 119

Qy        117 ------DNTPDGDFIIND--ASAVPIQF----------SNGSQSILLPNVIIMGAEPDLF 158

Db        120 PNIFTFGKTPKSNKKLNSLVTSTIPFPFGGYRETNYIESQNNKNFYASNIIIFGPGSNIV 179

Qy        159 ETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSL 218

Db        180 ENNV--IYYKKN--DAENGMGTMAEIVFQPLLTYKY-----NKFYIDPAMELTKCLIKSL 230

Qy        219 HGLYGAK---GITTKYTI-TQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQ---SNDI 271

Db        231 YFLYGIKPSDNLVVPYRLRTELDNKQFSQL---NIIDLLISGGVDLEFINTNPYWFTNSY 287

Qy        272 YTN---LLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIF 328

Db        288 FPNSIKMFEKYKNIYK--TEIEGNNAIGNDIKLRLKQKFQI--NVQDIWNLNLNYFCQSF 343

Qy        329 KKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN---NLKVNFR 385

Db        344 NSI-------IPDRFSNALKHFYRKQY-----------YTMDYTDNYNINGFVNGQINTK 385

Qy        386 GQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYND 445

Db        386 LPLSNKNTNIISK------PEKVV----NLVNENNIS-------------LMKSNIYGD 421

Qy        446 DNINTPKEIDDT--VTSNNNYE---NDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPK 500
```

```
Db        422 GLKGTTEDFYSTYKIPYNEEYEYRFNDSDNFPLN-------------NISIEEVDSIPE 467

Qy        501 ------YDSNG-----------TSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDT 543
                    |  |           | ::  |    :|   || ||      || | | | ||
Db        468 IIDINPYKDNSDNLVFTQITSMTEEVTTHTALSIN---YLQAQIT----NNENFTLSSDF 520

Qy        544 ALLEQPK----IYTFFSS-----EFINNVNKPVQA-ALFVSWIQQVLVDFTTEANQKSTV 593
                : :   |    :|:|  :      | | | : |:     :  |:::|  ::: : |    :
Db        521 SKVVSSKDKSLVYSFLDNLMSYLETIKN-DGPIDTDKKYYLWLKEVFKNYSFDINLTQEI 579

Qy        594 DKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIK-S 652
                | :  |: || : | |||| | :   :| :  :  || ' |:  :  |  | | :  |  |
Db        580 DSMCGINEVVLWFGKALNILNTS--NSFVEEYQDSGAISLISKKDNLREPNIEIDDISDS 637

Qy        653 FLGSSDNKNKVIKAINNALKERDEK----WKEVYSFIVSNWMTKINTQFNKRKEQMYQAL 708
                || |       | :|| | |    |     :|::|   :  | |:  :|: :      |::
Db        638 LLGLS------FKDLNNKLYEIYSKNIVYFKKIYFSFLDQWWTEYYSQYFELICMAKQSI 691

Qy        709 QNQVNALKAIIESKYNSY--------TLEEKNELTNKYDIEQIENELNQKVSIAMNNIDR 760
                | : :| |:::|:           ||:   | | |  |     :|: :  |:|| :|
Db        692 LAQESLVKQIVQNKFTDLSKASIPPDTLKLIRETTEKTFI-----DLSNESQISMNRVDN 746

Qy        761 FLTESS------------ISYLMKLINEVKINKLREY-------DENVKTYLLDYIIKHG 801
                || ::|            |||: | || : | | ||:       ::| |: |
Db        747 FLNKASICVFVEDIYPKFISYMEKYINNINI-KTREFIQRCTNINDNEKSIL-------- 797

Qy        802 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN----KFFKRIKSSSVLNMRY 857
                      :||    |    | ||   : | : : ::||    :  | | |   | :
Db        798 ---------INSYTFKT----IDFK---FLDIQSIKNFFNSQVEQVMKEILSPYQLLL-- 839

Qy        858 KNDKYVDTSGYDSNI--NINGDVYKYPTNKNQFGIYNDKLSEVN---------ISQNDYI 906
                    : | :|||  :|:|         ||    | : :  |             | |: |
Db        840 -----FASKGPNSNIIEDISG--------KNTLIQYTESIELVYGVNGESLYLKSPNETI 886

Qy        907 IYDNKY------KNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
                 : ||:       ||:| ||:|            |:: | :   : || ||::    | ::
Db        887 KFSNKFFTNGLTNNFTICFWLR-------FTGKNDDKTRLIGNKVNNCGWEIYFEDNGLV 939

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYIN-KWIFVTITNDRLGDSKL-YINGNLIDKKSIL 1018
                : : |::| ||:  :         :|: ||  | :::|: ||| |  | :||   :   ||
Db        940 FEIIDSNG-NQESVY-------LSNIINDNWYYISISVDRLKDQLLIFINDKNVANVSID 991

Qy       1019 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGN 1078
                : :| | ||:  :|| : : |   :   :: |   :   |:   | : : : ::|   :
Db        992 QILSIY-STNII-SLVNKNNSIY--VEELSVLDNPITSEEVIRNYFSYLDNSYIRDSSKS 1047

Qy       1079 YLLYDKEYYLLNVLKPNNF---INRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNS 1135
                | |:| | | | : |       :|    | ||: |      :       | :| |:  :: |
Db       1048 LLEYNKNYQLYNYVFPETSLYEVNDNNKSYLSLKNTDGINI------SSVKFKLINIDES 1101

Qy       1136 STNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQV 1187
                    |:| |:  |                    || : ||   ||   |
Db       1102 KV---YVQKWDECII---------------CVLDGTEKYLDISPENNRIQLV 1135


RESULT 12
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;

  Query Match             11.3%;  Score 735;  DB 1;  Length 1162;
  Best Local Similarity   23.4%;  Pred. No. 1e-22;
  Matches  302;  Conservative  248;  Mismatches  484;  Indels  254;  Gaps   58;

Qy          6 SFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSLK-NG 64
              : | : ||:|: :  ::     | |:|:| :  ||||:|||              ||| |
Db          6 NLNIDSPVDNKNVAIVRSRN-QMFFKAFQVAPNIWIVPERYY----------GESLKINE 54

Qy         65 DSSY----YDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
               |  :    || |:| :: ||| ||:   |:  |||:|: |  ||  :| | |:   :||
Db         55 DQKFDGGIYDSNFLSTNNEKDDFLQATIKLLQRINNNVVGAKLLSLISTAIPFPYENNT- 113

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
                        |          |  ::      |::| |   :: : |   |  :  |  | |:
Db        114 ------EDYRQTNYLSSKNNEHYYTANLVIFGPGSNIIKNNV--IYYKKEYAES--GMGT 163

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAK---GITTKYTITQKQ 237
              :  : | |      |  :  :||  |||| |:  || ||: ||| |    :   | :  :
Db        164 MLEIWFQP-----FLTHKYDEFYVDPALELIKCLIKSLYYLYGIKPNDNLNIPYRLRNEF 218

Qy        238 NPLITNIRGTNIEEFLTFGGTDLNI-------------ITSAQSNDIYTNLLADYKKIAS 284
              | |        |: :||  || |  :               | :::: : | |   ||:
Db        219 NSL--EYSELNMIDFLISGGIDYKLLNTNPYWFIDKYFIDTSKNFEKYKN---DYE---- 269

Qy        285 KLSKVQVSNPLLNPYKDVFEAKYGLD-KDASGIYSVNINKFNDIFKKLYSFTEFDLATKF 343
                |:: :| : |  |   | |: :: ||   |: :|:: |                : :|
Db        270 --IKIKNNNYIANSIKLYLEQKFKINVKD---IWELNLSYF--------------SKEF 309

Qy        344 QVKCRQTYIGQYKYFKLSNLLND------SIYNISEGYNINNLKVNFRGQNANLNPRIIT 397
              |:   : |          :| ||        : :  : ||||  |           | :| |
Db        310 QIMMPERY---------NNALNHYYRKEFYVIDYFKNYNINGFK----------NGQIKT 350

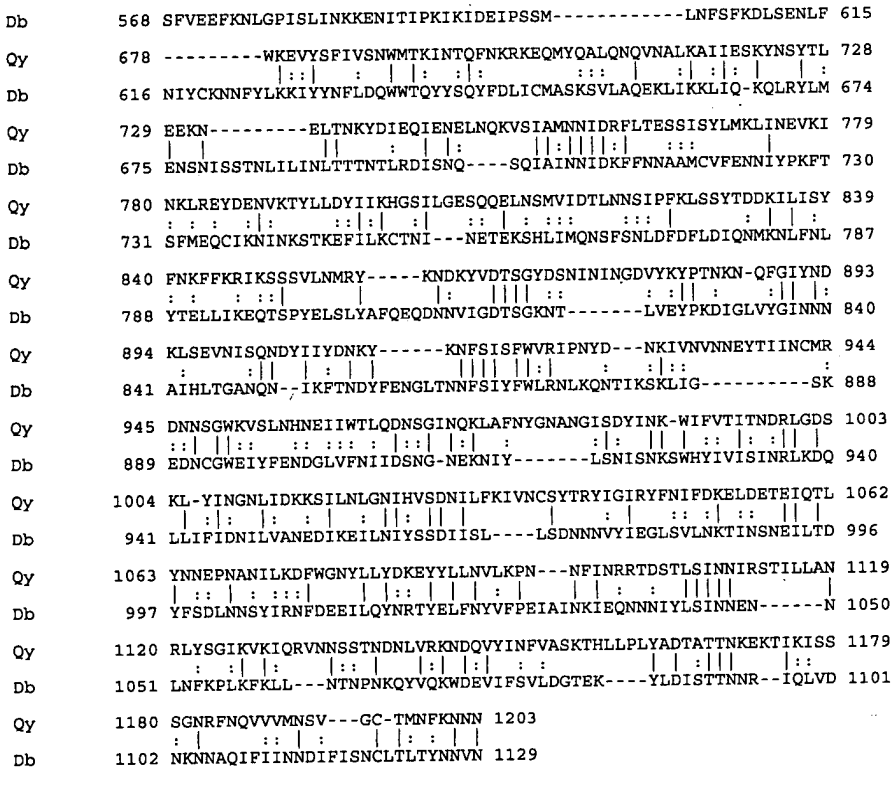
Qy        398 PITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYND------DNINTP 451
               :      |:||   : ||::         |  ||  |       | | |       ||  :
Db        351 KLPLSKYNKEIINKPELIVNL--------INQNNTVLM---KSNIYGDGLKGTVDNFYS- 398

Qy        452 KEIDDTVTSNNNYENDLDQVIL-NFN---SESAPGLSDEKLNLTIQN-DAYIPKYDSNGT 506
                :  :  | |||: ::   | | |     |  | ::|| :     :| | :|| |:
Db        399 ---NYIIPYNLNYEHSINYFYLDNVNIEEIEKIPPINDEDIYPYRKNADTFIPVYNITKA 455

Qy        507 SDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPK-IYTFFSS-----EFI 560
               :|       :|   || || :    |::||:|     :  :   :|:| ::      |||
Db        456 KEINTTTPLPVN---YLQAQMI--DSNDINLSSDFLKVISSKGSLVYSFLNNTMDYLEFI 510

Qy        561 NNVNKPVQA-ALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKG 619
              :||:     :  |:: :  ::: :   :       : ::|:|| |||| |
Db        511 -KYDKPIDTDKKYYKWLKAIFRNYSLDITETQEISNQFGDTKIIPWIGRALNILN--TNN 567

Qy        620 NFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEK-- 677
              :| :  : ||   |:  :  : || | :  | | :            :| : |:  |
```



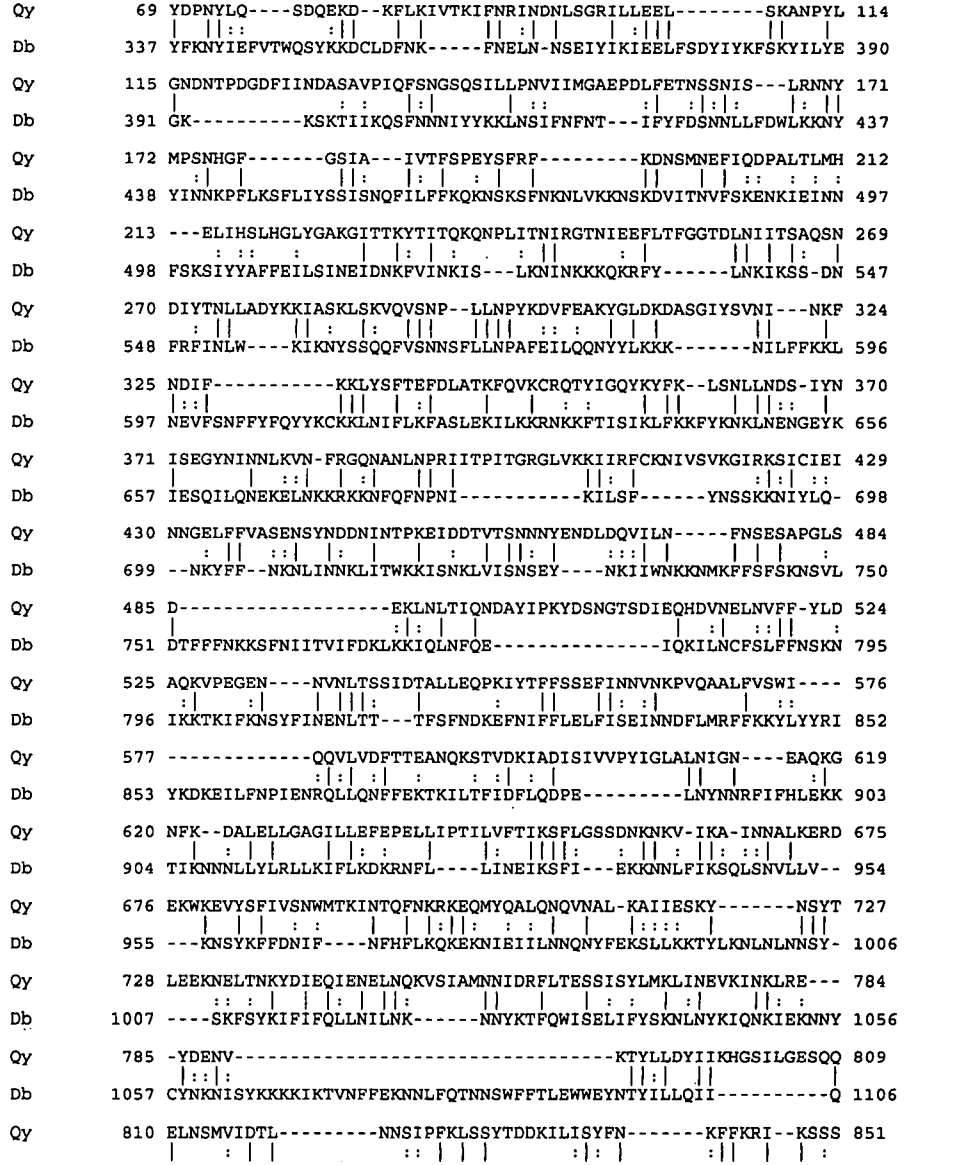
```
RESULT 13
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match           11.2%;  Score 727.5;  DB 1;  Length 1162;
  Best Local Similarity  22.5%;  Pred. No. 2e-22;
  Matches  298;  Conservative  256;  Mismatches  508;  Indels  261;  Gaps   56;
```



```
RESULT 14
YCF2_MARPO
ID   YCF2_MARPO     STANDARD;      PRT;  2136 AA.
AC   P09975;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN   YCF2.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC   Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
CC   -!- SIMILARITY: Belongs to the ycf2 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04465; CAA28078.1; -.
DR   PIR; S01591; A05037.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
KW   Chloroplast; Hypothetical protein.
SQ   SEQUENCE   2136 AA;  259911 MW;  5BD170C9CCF61197 CRC64;

  Query Match             4.1%;  Score 269;  DB 1;  Length 2136;
  Best Local Similarity   20.6%;  Pred. No. 0.0013;
  Matches  308;  Conservative  208;  Mismatches  484;  Indels  498;  Gaps   78;
```



```
Db       1107 ETFFQITDVLEYFKKKKIIEKNLKFFLKSKKISLKTLSFHNFKLKWNLRFFNEINYKKNY 1166

Qy        852 VLNMRYKNDKYVDTSG----------------YDSNINI-----------NGDVYKYPTN 884
              :|| : :   ::                     |    :|           | :: :| |:
Db       1167 LLNFLWSDFNLINNCNNLYWVIFSLVIFIFLYYQKIFSIIIGSDCFHLWKNFEIIQYLTD 1226

Qy        885 KNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNY----------------- 926
              :::  :|  ||:  | :        || :|    |:  : :|
Db       1227 RSR-SLYFTKLTRRNKTAL------NKTENLLSYFFQNLTHYITNIKFYLLTKKNLKKWL 1279

Qy        927 -DNKIVNVNNE------YTIINCMRDNNSGWKVSLNHNEIIWT------LQDNSGI--NQ 971
              :|| ::::          ::|   :  | |::  || |:  :|      : :  |:   |
Db       1280 INNKTLDLSRRKRKLLVQSLITHNKIQNYGFE--LNSNKQFFTSYFGYQITNQQGLLYFQ 1337

Qy        972 KLA--FNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN---------------LIDK 1014
              ||  |      |   |  ||||  :  :       ||     |                |
Db       1338 YLAQFFQKNLINNSLDLANKWIVFSFWHKIFSSQKLRQTNNIELGFQNIPVPLQFGLSYS 1397

Qy       1015 KSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKD 1074
              | || :| |    : | |  | :  |:           | :  | |  |:|  ::: :
Db       1398 KGILLIGPIETGRSYLIK--NLAAESYV----------PLFKISINKLLYNKP--DVITE 1443

Qy       1075 FWGNYLL-----------YDKEY--------------------------YLLNVL---- 1092
              | | |:           : |:                          :|| :|
Db       1444 SWMNILIESLRRLNLTLDFAKKMSPCIIWIQNIHQLNVNRLTQNVESDPTFLLGILLKYF 1503

Qy       1093 -------KPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLV--R 1143
                    | || |     ||     :   ::  |||   | |::  ::       |:  :
Db       1504 QTDFSKTKKNNII--VIGSTHLPKKVDPALISPNRLDKIINVRLFNISQRKKQFPLLLKK 1561

Qy       1144 KNDQV-----YINFVASKTHLLPLYADTATTNKEKTIKISSSGN---------------- 1182
              || |:     ::|   |:|    |   :| ||:   | |: : :
Db       1562 KNFQLKENLFFLNEFGSRTMGYNLRDLSALTNEVLLISITKNRSFIDTDTLKLAFHRQIF 1621

Qy       1183 ---------RFNQV--VVMNSVGCT----MNFKNNNGN--NIGLLGFKADTVVASTWY 1223
                      |::  :|:  || |    :  |::: |  |||   :| :    | ||
Db       1622 GLTYTNNKLNFDRIFKIVIYKVGKTIIQNILIKSSSMNLLNIGNFLWKKNFYYLSKWY 1679
```

```
RESULT 15
ATX1_PLAFA
ID   ATX1_PLAFA     STANDARD;      PRT;  1956 AA.
AC   Q04956;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cation-transporting ATPase 1 (EC 3.6.3.-).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T9/96;
RX   MEDLINE=93132070; PubMed=8421054;
RA   Krishna S., Cowan G., Meade J.C., Wells R.A., Stringer J.R.,
RA   Robson K.J.;
RT   "A family of cation ATPase-like molecules from Plasmodium
RT   falciparum.";
RL   J. Cell Biol. 120:385-398(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC       ATPases). Subfamily V.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65738; CAA46646.1; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR008250; E1-E2_ATPase_reg.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   TIGRFAMs; TIGR01494; ATPase_P-type; 7.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
FT   DOMAIN       1     35       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    36     58       POTENTIAL.
FT   DOMAIN      59     61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    62     80       POTENTIAL.
FT   DOMAIN      81    407       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   408    427       POTENTIAL.
FT   DOMAIN     428    440       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   441    462       POTENTIAL.
FT   DOMAIN     463   1818       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1819   1837       POTENTIAL.
FT   DOMAIN    1838   1845       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1846   1863       POTENTIAL.
FT   DOMAIN    1864   1881       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1882   1905       POTENTIAL.
FT   DOMAIN    1906   1928       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1929   1952       POTENTIAL.
FT   DOMAIN    1953   1956       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES    496    496       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL     1760   1760       MAGNESIUM (BY SIMILARITY).
FT   METAL     1764   1764       MAGNESIUM (BY SIMILARITY).
FT   DOMAIN     246    251       POLY-ASN.
FT   DOMAIN     252    256       POLY-LYS.
FT   DOMAIN     937    941       POLY-ASN.
FT   DOMAIN    1344   1347       POLY-LYS.
FT   DOMAIN    1363   1372       POLY-ASN.
FT   DOMAIN    1680   1684       POLY-ASN.
SQ   SEQUENCE   1956 AA;  230285 MW;  AE708AAE99009335 CRC64;
```

Query Match 4.1%; Score 266.5; DB 1; Length 1956;
Best Local Similarity 20.4%; Pred. No. 0.0014;
Matches 281; Conservative 201; Mismatches 483; Indels 415; Gaps 67;

```
Qy          4 INSFNY--NDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 61
              ||: |:     :|:  :||    |   |    | ||: :   :||:
Db        257 INNLNFVKGTYINSNDLLYDDKIGVNIFEDDVNNMKHKF--NQRNI-------------- 300

Qy         62 KNGDSSYY--DPNYLQSDQEK----DKFLKIVTKIFNRINDNLSGRILL--EELSKANPY 113
                 :||  | | |: : :     |  || |  |  :       :|:   |:::|  |
Db        301 -----NYYNKDTNNLEYNNKHRYIYDCLLKKVEAISQK------NKIIYSNEDINKYMLY 349

Qy        114 LGNDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRN---- 169
                  | ::::  :   |:::|  :: :|  ||  |     | |    | : |
Db        350 -------GGTYVLSLYNINKIKYNNKEENRILGLVIKTG-----FITTKGKI-VNNILYH 396

Qy        170 -----NYMPSNHGFGSIAIV--TFSPEYSFRFKDNSMNEF-----------IQD---PAL 208
                  | :  :: |  | |:    ||  :   :   | ||:           | |   |||
Db        397 KKKELNLINDSYKFLIILIIYALFS-VFILLYITLSNNEYTNHIIIKCLDIITDAIPPAL 455

Qy        209 --TLMHELIHSLHGLYGAKGITT----KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNI 262
                ||   :  ::  |     |:     |  | : | :: :  ||  |  | |    : |
Db        456 PTTLTVGISIAISRLKKKFSISCLCPHKINIAGQINTMVFDKTGTLTENNLQF----IGI 511

Qy        263 ITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVN-- 320
              ||  : |     |:|:|:  |                   |::    |   || : |:: |
Db        512 ITQNKKN---KNMLSDFIHI----------------KEMNTESYIHSKDDNMIHNKNSI 551

Qy        321 INKF--NDIFKKLYSFTEFDLATK----FQVK------CRQTYIGQYK--YFKLSNLLND 366
              |:::   |  | |:: ::    ||     | |:       :  || : |  |:  :   ||
Db        552 ISEYYIKDNMKNLHTSSKKKSITKERSNFLVQTIKSCLLKDHYIKEKKKEYYTNNTYCND 611

Qy        367 SIYNIS----------------EGYNINNL-KVNFRGQNANLNPRIITPITGRGLVKKII 409
                | |                | |||::|  :| :  : |    ::   :   |: | |
Db        612 LHINDSTCSSYLLNSETKDAYCEYYNIDHLCDINKKNMDINSKNELMGKYSKNELMGKTI 671

Qy        410 --------------RFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEI 454
                           :: || :  |  :  :  : :  ||     |:|       |     :
Db        672 KNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKTIKNQVGV 731
```

```
Qy       455 DDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSD----IE 510
Db       732 DTNI-YHMNCDND------------------------YNYDYPCDYNCNNCNDTYHRLE 765

Qy       511 QHDVNELNVFFYLDAQKVP----EGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKP 566
Db       766 YHNINKDNSF------NIPPEKNKSYNNISEHIKINYPLLFEALACCHTLSKVNNKIMGD 819

Qy       567 VQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALE 626
Db       820 VLEILMFNFTN---CDMLINNNSFIIKEKKKNCSYDFQKIDGDKNIGANDERCHLNN--N 874

Qy       627 LLGAGILLEFEPELLIPTILVFTIKSFLGSS------------DNKNKVIKAINNALKE 673
Db       875 LVSYNILKRFEFQSRLQRMSVI-VKSTYGNNNDDNNDDDNNNDDDNNDDNNNDDNNNDDN 933

Qy       674 RDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNE 733
Db       934 NDDNNNNNYYYNI----------FCKGSPEKIKEL-----CLKSKIPNNY--------DE 970

Qy       734 LTNKYD-----IEQIENELNQKVSIAMNNIDRFLTESSISYLMKLI--NEVKINK----- 781
Db       971 ILNKYTKQGMRILSISYKRVKSKNINLLNVKRSFVESNLHFLGFLIFTNNMKKNAPDIIH 1030

Qy       782 ---------LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTL------NNSIPFK 826
Db      1031 NLQTSGCQCIMSTGDNVLTSI--HVAKKCGIINSN---VESIIIGDVIPVVGKNNKQKKK 1085

Qy       827 LSSYT----------DDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNI---- 872
Db      1086 LVWYNHKNDTYLKGHDKTCIDNEFTSIQSQMNSDNIC-----GDNICGDNIYGDNICGDN 1140

Qy       873 ----NINGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKY-KNFSISFWVRIP--- 924
Db      1141 IYGDNINGD------NINGDNIYGDNINGDNINGDNINTYDNIYGDNYNLDY---CPTEY 1191

Qy       925 ---NYDNKIVNVNN----------EYTIINCMRDNNSGWKVSLNHNEI-------IWTL 963
Db      1192 HKCTYNNSILYRNNFLYKKENKKDKNYKNISTLYEHRT--------NDIQFDKLCDILIN 1243

Qy       964 QDNSGINQKL---AF------------NYGNANGISDYI----NKWIFVTITNDRLGDS 1003
Db      1244 KDPRNVNIVLTGKAFIFLKKKFYSFHLPYYEECKNIVHYIMKKKHKKIKNIINNH---NS 1300

Qy      1004 KLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLY 1063
Db      1301 NLYYHYNIID--------------TFVKRMNKEYM---------CFNKLLYKIQQKLLY 1336

Qy      1064 NNEPNANILKDFWGNYLLYD-KEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLY 1122
Db      1337 NLIHNLYKKKKYMNNY--YDIDEVHLIGNNNNNNNKNNSKEKKLPLKN------------ 1382

Qy      1123 SGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLP---------LYADTATTNKEK 1173
Db      1383 -----KMKHIRKNESNDNI------TFNTYTSNNIHLSKYKYVHHKNYYYPDSCTNLRKK 1431
```

Search completed: September 10, 2004, 03:35:27
Job time : 24.2517 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 57.1937 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-50
Perfect score: 6511
Sequence: 1 MPTINSFNYNDPVNNRTILY.....TNSNGFFWNFISEEHGWQEK 1250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6496	99.8	1252	2	Q8KZM3	Q8kzm3 clostridium
2	6496	99.8	1255	2	Q9FAR6	Q9far6 clostridium
3	6171.5	94.8	1251	2	Q9K395	Q9k395 clostridium
4	4216	64.8	1278	2	Q57236	Q57236 clostridium
5	4145.5	63.7	1268	2	Q45851	Q45851 clostridium
6	4024.5	61.8	1280	2	Q9ZAJ5	Q9zaj5 clostridium
7	2228	34.2	1291	2	Q8GR96	Q8gr96 clostridium
8	2207	33.9	1291	2	Q933K0	Q933k0 clostridium
9	2203	33.8	1291	2	Q9ZAJ8	Q9zaj8 clostridium
10	2200.5	33.8	1291	2	Q93G71	Q93g71 clostridium
11	2181	33.5	1291	2	Q08077	Q08077 clostridium
12	2111	32.4	1310	2	Q93N27	Q93n27 clostridium
13	1875	28.8	367	2	Q45862	Q45862 clostridium
14	1866	28.7	367	2	Q45861	Q45861 clostridium
15	1860	28.6	1275	12	Q9QTG7	Q9qtg7 clostridium
16	1852	28.4	1285	2	Q45967	Q45967 clostridium
17	1849	28.4	1285	2	Q9LBR1	Q9lbr1 clostridium
18	1815	27.9	1291	2	Q93HT3	Q93ht3 clostridium
19	1769.5	27.2	1280	2	Q9LBS7	Q9lbs7 clostridium
20	1761.5	27.1	1280	2	Q45849	Q45849 clostridium
21	1760.5	27.0	1280	2	Q841S3	Q841s3 clostridium
22	975.5	15.0	347	2	Q840G4	Q840g4 clostridium
23	880	13.5	173	2	Q9KW90	Q9kw90 clostridium
24	867	13.3	173	2	Q9KW88	Q9kw88 clostridium
25	865	13.3	173	2	Q9KW89	Q9kw89 clostridium
26	858	13.2	173	2	Q9K345	Q9k345 clostridium
27	784	12.0	361	2	Q45848	Q45848 clostridium
28	784	12.0	361	2	Q45846	Q45846 clostridium
29	770	11.8	1196	2	Q53550	Q53550 clostridium
30	768	11.8	1160	2	O87710	O87710 clostridium
31	766	11.8	1196	2	Q45916	Q45916 clostridium
32	766	11.8	1196	9	Q38197	Q38197 clostridium
33	764	11.7	1196	2	Q93HT4	Q93ht4 clostridium
34	763	11.7	1163	2	Q45850	Q45850 clostridium
35	760	11.7	1196	9	Q9ZX77	Q9zx77 clostridium
36	753.5	11.6	1198	2	O06018	O06018 clostridium
37	753	11.6	1196	2	Q9LBS8	Q9lbs8 clostridium
38	747	11.5	1193	2	P71108	P71108 clostridium
39	747	11.5	1196	2	Q9LBR2	Q9lbr2 clostridium
40	727	11.2	1159	2	P71109	P71109 clostridium
41	722.5	11.1	1162	2	Q9ZAJ6	Q9zaj6 clostridium
42	716.5	11.0	441	2	Q9X708	Q9x708 clostridium
43	712	10.9	1197	2	Q45888	Q45888 clostridium
44	705.5	10.8	1161	2	O69276	O69276 clostridium
45	696.5	10.7	1193	2	P71107	P71107 clostridium

ALIGNMENTS

RESULT 1
Q8KZM3
ID Q8KZM3 PRELIMINARY; PRT; 1252 AA.
AC Q8KZM3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 5262;
RA Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA Nakamura S., Karasawa T., Kozaki S.;
RT "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT complete cds.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB088207; BAC05434.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1252 AA; 143510 MW; 41B633BB744D3B41 CRC64;

Query Match 99.8%; Score 6496; DB 2; Length 1252;

```
  Best Local Similarity   99.8%;  Pred. No. 2.2e-242;
  Matches 1249;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1252


RESULT 2
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match             99.8%;  Score 6496;  DB 2;  Length 1255;
  Best Local Similarity   99.8%;  Pred. No. 2.2e-242;
  Matches 1249;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 63

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 123

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 183

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        184 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 243

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 303

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 363

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 423

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 483

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 543

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        544 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 603

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 663

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 723

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 783

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 843

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 903

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 963

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1023

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1083

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1143

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              |||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db       1144 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1203

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1255


RESULT 3
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
```

```
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             94.8%;  Score 6171.5;  DB 2;  Length 1251;
  Best Local Similarity   95.0%;  Pred. No. 6.9e-230;
  Matches 1189;  Conservative   23;  Mismatches   37;  Indels    3;  Gaps    2;

Qy          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60
              || |||||||||||:|||||||||||:|||||||||||||||||||| |||| ||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120
              |||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFVPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              |||||||||  || | ||||||||||||||||||||||||||||||||||||||||||||
Db        721 FKYNSYTLEEKKELKNNYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIQHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              |||||||||||||||||||||||||||||||||||||||||:::  |||||||  |:| | |||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFNSKPSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNINNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020
              ||||||: ||||| | ||||||||||||||||||||||||||||||||:|||:|||||||
Db        961 WTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGHLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080
              |||||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              |||| ||||||||||||:|| ||||||||||||||||||||||||||||:|||||
Db       1081 LYDKGYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNDSSTNDR 1140

Qy       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198
              ||||||||:: | :    ||||| ||:|||||| ||||||||||||||  ||||
Db       1141 FVRKNDQVYINYI-SNSSSYSLYADTNTTDKEKTIKSSSSGNRFNQVVVMNSVGNNCTMN 1199

Qy       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1200 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1251


RESULT 4
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           64.8%;  Score 4216;  DB 2;  Length 1278;
  Best Local Similarity  63.1%;  Pred. No. 1.3e-154;
  Matches  809;  Conservative  210;  Mismatches  225;  Indels   38;  Gaps   16;

Qy          1 MP-TINSFNYNDPVNNRTILYIK---PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              ||  |||||||||||: ||||::       :::||:| ||:|:|||||| ||| | ||
Db          1 MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              || ||:|| |:|||||| :| |||::||    |:| ||| | :| :||:|:|| | |||||
Db         61 PPASLENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGN 120

Qy        117 DNTPDGDF-IINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNI-SLRNN---Y 171
              ::||  :|  :   ::| |: |   :| :: |::::|| ||:|| :|  :  | ::   |
Db        121 EHTPINEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVY 180

Qy        172 MPSNHGFGSIAIVTFSPEYSFRFKD------NSMNEFIQDPALTLMHELIHSLHGLYGAK 225
              ||| ||||| |||||||| : | |      :|   || |||::| |||||:|||||||:
Db        181 DPSNDGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGAR 240

Qy        226 GITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              |:| | ||  || ||:   :   :|||||||| ||||||||    || ||||:|:|||::
Db        241 GVTYKETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATR 300

Qy        286 LSKVQVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKF 343
              ||:|  : |   :| ||| |: |||||||:| | |:|| ||||:|:|||||||| ||| ||
Db        301 LSRVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKF 360

Qy        344 QVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRG 403
              :|||| ||  :| : |: |||:| || :|||:|| || || |||  |||:||  |  :|
Db        361 KVKCRNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKG 420

Qy        404 LVKKIIRFCKNIVSVKGIR--KSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSN 461
              ||:||::|||:::  || :   :|| :|| ||||||||:|||:::|||||||||   |
Db        421 LVEKIVKFCKSVIPRKGTKAPPRLCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLN 480

Qy        462 NNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFF 521
              ||| |:||:|||::|||: | :|:: ||  :|:|:|:|:|||||||:||:|:| :|||||
Db        481 NNYRNNLDEVILDYNSETIPQISNQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFF 540

Qy        522 YLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLV 581
              || |||||||| |::|||||||||| |: ::|||||||||| :|||| ||||:||| ||:
Db        541 YLHAQKVPEGETNISLTSSIDTALSEESQVYTFFSSEFINTINKPVHAALFISWINQVIR 600

Qy        582 DFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELL 641
              |||||| |||| |||||||:||||:||||||||| || |||:| ||||||||||| ||||
Db        601 DFTTEATQKSTFDKIADISLVVPYVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELL 660

Qy        642 IPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRK 701
              |||||||||||:|||:||||:|||||:| ||: ||||:||:|||||:|:|||||||||
Db        661 IPTILVFTIKSFIGSSENKNKIIKAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRK 720

Qy        702 EQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRF 761
              ||||||||||:|:| :|| |||:|| :|:|  | ::|:|   |  |||:|||:|| ||:||
Db        721 EQMYQALQNQVDAIKTVIEYKYNNYTSDERNRLESEYNINNIREELNKKVSLAMENIERF 780

Qy        762 LTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNN 821
              :||||| |||||||| |::|||||||| || |||||| :| |||| | |||| :|  ||||
Db        781 ITESSIFYLMKLINEAKVSKLREYDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNN 840

Qy        822 SIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKY 881
              ||||:|||||:||||| |||| :|:|| :|:|:|||:|:|::| ||| |||:|||||| |
Db        841 SIPFELSSYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIY 900

Qy        882 PTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIIN 941
              ||:|||||: | |||||:||: |||: :|:|||||||||||| | || ||:|||||||:
Db        901 STNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIID 959

Qy        942 CMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLG 1001
              |:|:||||||:|||:|:|||||| :| |||| |||     |||||||||||||:|||
Db        960 CIRNNNSGWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLG 1019

Qy       1002 DSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQT 1061
              :|::|||||||:||| |||:|||||||||||||| |: |||:||||| :|| || :|||:|
Db       1020 NSRIYINGNLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIET 1079

Qy       1062 LYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRS-----TIL 1116
              ||::||: :||||||||||||:| |||||:|: :  | : : : |:||  |       |
Db       1080 LYSDEPDPSILKDFWGNYLLYNKRYYLLNLLRTDKSITQNS-NFLNINQQRGVYQKPNIF 1138

Qy       1117 LANRLYSGIKVKIQRVNNSSTN----DNLVRKNDQVYINFVASKTHLLPLYADTATTNKE 1172
                 |||:|::| |::  | ||:    || |||||  ||| |  :     |||| :   |
Db       1139 SNTRLYTGVEVIIRK--NGSTDISNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPE 1195

Qy       1173 KTIKI---SSSGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHM 1227
              | ||:   |:| |   |::||:|:|  |||||:|||| ||||||| :: :|||:||| ::
Db       1196 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 1255

Qy       1228 RDNTNSNGFFWNFISEEHGWQE 1249
              | ||:||| ||:|||:||||||
Db       1256 RKNTSSNGCFWSFISKEHGWQE 1277


RESULT 5
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match             63.7%;  Score 4145.5;  DB 2;  Length 1268;
  Best Local Similarity   62.5%;  Pred. No. 6.9e-152;
  Matches  795;  Conservative  221;  Mismatches  229;  Indels   27;  Gaps   15;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              ||  ||:||||||:|| ||||:|        ::||:| || |:|||||||:||  | ||
Db          1 MPVNINNFNYNDPINNTTILYMKMPYYEDSNKYYKAFEIMDNVWIIPERNIIGKKPSDFY 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              || || :| |:|||||| :| |||:||| | |:||||| | :|::||||:     |||||
Db         61 PPISLDSGSSAYYDPNYLTTDAEKDRFLKTVIKLFNRINSNPAGQVLLEEIKNGKPYLGN 120

Qy        117 DNTPDGDFIINDAS-AVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNN---YM 172
              |:|   :|  |: | :| |: |||:   :| |::|:|  |:: | ::  : :  |   |
Db        121 DHTAVNEFCANNRSTSVEIKESNGTTDSMLLNLVILGPGPNILECSTFPVRIFPNNIAYD 180

Qy        173 PSNHGFGSIAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYT 232
              ||  ||||| :::|| || : | ||: : || |||::| ||||| |||||||||:| |
Db        181 PSEKGFGSIQLMSFSTEYEYAFNDNT-DLFIADPAISLAHELIHVLHGLYGAKGVTNKKV 239

Qy        233 ITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVS 292
              |   |  |:  :   ||||:|||| |||||:: :  ||  ||::|  |||:||:|  :
Db        240 IEVDQGALMAAEKDIKIEEFITFGGQDLNIITNSTNQKIYVILLSNYTAIASRLSQVNRN 299

Qy        293 NPLLNP--YKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQT 350
              |  ||   ||: |: |||||:|::| |:|||:||| |:|||:|||| ||| ||||| |
Db        300 NSALNTTYYKNFFQWKYGLDQDSNGNYTVNISKFNAIYKKLFSFTECDLAQKFQVKNRSN 359

Qy        351 YIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIR 410
              |:  :| |:| :||:|:||:||||:|| :|:||  ||| ||| ||: ||   |||:: :
Db        360 YLFHFKPFRLLDLLDDNIYSISEGFNIGSLRVNNNGQNINLNSRIVGPIPDNGLVERFVG 419

Qy        411 FCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQ 470
              ||:||| || : |:||::|| :|||||||:|||:: ||:|||||  :||||: :||:
Db        420 LCKSIVSKKGTKNSLCIKVNNRDLFFVASESSYNENGINSPKEIDDTTITNNNYKKNLDE 479

Qy        471 VILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPE 530
              |||::||:: | ||   || | |||:|:|||||||||:|::: |::|||||| ||| ||
Db        480 VILDYNSDAIPNLSSRLLNTTAQNDSYVPKYDSNGTSEIKEYTVDKLNVFFYLYAQKAPE 539

Qy        531 GENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQK 590
              ||: ::||||::||||:  |:|||||:||| |||||||||:|||||: |||||| ||
Db        540 GESAISLTSSVNTALLDASKVYTFFSSDFINTVNKPVQAALFISWIQQVINDFTTEATQK 599

Qy        591 STVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTI 650
              ||:|||||||::|||:|||||||| ||||||:|:|||||||||| ||||||||||||
Db        600 STIDKIADISLIVPYVGLALNIGNEVQKGNFKEAIELLGAGILLEFVPELLIPTILVFTI 659

Qy        651 KSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQN 710
              |||: | |:|||:|||||||||:|||: |||||||:|||||:|:|||||||||||||||||
Db        660 KSFINSDDSKNKIIKAINNALRERELKWKEVYSWIVSNWLTRINTQFNKRKEQMYQALQN 719

Qy        711 QVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYL 770
              ||: :| ||| |||:|||:||| |  :|:|  |: |||:|||:|| ||||||||||||||
Db        720 QVDGIKKIIEYKYNNYTLDEKNRLRAEYNIYSIKEELNKKVSLAMQNIDRFLTESSISYL 779

Qy        771 MKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQ-QELNSMVIDTLNNSIPFKLSS 829
              |||||| ||||| |||: |  |||:||::: | || |    |||:|:| :|||||||:||
Db        780 MKLINEAKINKLSEYDKRVNQYLLNYILENSSTLGTSSVPELNNLVSNTLNNSIPFELSE 839

Qy        830 YTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFG 889
              ||:|||||    :|:|||  ||:|||:|:|::::|:||| |||:||||:| | ||:||||
Db        840 YTNDKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFG 899

Qy        890 IYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSG 949
              ||: :|||||:||: |||:::|:|||:|||||| |:| : |:|||||||||||:||||
Db        900 IYSSRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN-LKNLNNEYTIINCMRNNNSG 958

Qy        950 WKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYING 1009
              ||:|||:| ||||||| :| |||| |||      |||||||| |||||:||| |||||||
Db        959 WKISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYING 1018

Qy       1010 NLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069
              || |:|||||||||| |||||||| |: |||:|||| ||: |||:|||:|||::||::
Db       1019 NLTDQKSILNLGNIHVDDNILFKIVGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR----STILLAN-RLYSG 1124
              |||||||||||:|:||||:|||| : : :|  |:||  |     | : :| |||:|
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPNMSVTKNSD-ILNINRQRGIYSKTNIFSNARLYTG 1137

Qy       1125 IKVKIQRVNNSSTN--DNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIK---ISS 1179
              ::| |::| :: |:  || ||||| |||| |   :    |||| :|: |||||   ||:
Db       1138 VEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSE-YQLYADVSTSAVEKTIKLRRISN 1196

Qy       1180 SGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFF 1237
              |    ||:::|:|:|  |||||| ||||:|||||  : :|||:||| ::|:|| :|| |
Db       1197 SNYNSNQMIIMDSIGDNCTMNFKTNNGNDIGLLGFHLNNLVASSWYYKNIRNNTRNNGCF 1256

Qy       1238 WNFISEEHGWQE 1249
              |:||||:||||||
Db       1257 WSFISKEHGWQE 1268


RESULT 6
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
```

```
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             61.8%;  Score 4024.5;  DB 2;  Length 1280;
  Best Local Similarity   60.4%;  Pred. No. 3.1e-147;
  Matches  776;  Conservative  226;  Mismatches  243;  Indels   39;  Gaps   17;

Qy          1 MP-TINSFNYNDPVNNRTILYI-KP--GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              ||  |||||||||||: ||||: ||     :::||:| || |:||:|||: ||| | :|
Db          1 MPVVINSFNYNDPVNDETILYMQKPYEERSRKYYKAFEIMPNVWIMPERDTIGTKPDEFQ 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGN 116
              | |||||  |:|||||| :| |||::|| : |:|||||  | :|::|||||:| | ||||:
Db         61 VPDSLKNGSSAYYDPNYLTTDAEKDRYLKTMIKLFNRINSNPTGKVLLEEVSNARPYLGD 120

Qy        117 DNTPDGDFI-INDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNN----Y 171
              |:|  .:|: :|  ::| |:||   :| :: |::::|| ||:|:   : :   |
Db        121 DDTLINEFLPVNVTTSVNIKFSTDVESSIISNLLVLGAGPDIFKAYCTPLVRFNKSDKLI 180

Qy        172 MPSNHGFGSIAIVTFSPEYSFRFKD------NSMNEFIQDPALTLMHELIHSLHGLYGAK 225
              |||||||||  |:|||||    | |       ||   || |||::| |||||:||||||||
Db        181 EPSNHGFGSINILTFSPEYEHIFNDISGGNHNSTESFIADPAISLAHELIHALHGLYGAK 240

Qy        226 GITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              :| | ::  :: ||:    :   :|||||||| ||||| ||    || :|||:|:|||::
Db        241 AVTHKESLVAERGPLMIAEKPIRLEEFLTFGGEDLNIIPSAMKEKIYNDLLANYEKIATR 300

Qy        286 LSKVQVSNP--LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKF 343
              | :|  : |   :| ||| |: |||||::| | |:|| ||||:|:|||||||| ||| ||
Db        301 LREVNTAPPGYDINEYKDYFQWKYGLDRNADGSYTVNRNKFNEIYKKLYSFTEIDLANKF 360

Qy        344 QVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRG 403
              :|||| ||  :| : |: |||:| || :|||:|| || || |||| ||||:||  |  :|
Db        361 KVKCRNTYFIKYGFVKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNINLNPKIIDSIPDKG 420

Qy        404 LVKKIIRFCKNIVSVKGIRK--SICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSN 461
              ||:|||:|||:|:  || ::  |:|| :|| ||||||||:|||: :|||||||||    |
Db        421 LVEKIIKFCKSIIPRKGTKQSPSLCIRVNNRELFFVASESSYNESDINTPKEIDDTTNLN 480

Qy        462 NNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFF 521
              ||| |:||:|||::|||: | :|:  ||  :|:::|:|:|||||||:||::|| : ||||
Db        481 NNYRNNLDEVILDYNSETIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYDVVDFNVFF 540

Qy        522 YLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLV 581
              || |||||||| |::||||||||||: |:|||||||||: :|||| ||||: || :|:
Db        541 YLHAQKVPEGETNISLTSSIDTALLEESKVYTFFSSEFIDTINKPVNAALFIDWISKVIR 600

Qy        582 DFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELL 641
              |||||| |||||||||||::|||:|||||   ||:||||::| ||||||||||| |||
Db        601 DFTTEATQKSTVDKIADISLIVPYVGLALNIVIEAEKGNFEEAFELLGAGILLEFVPELT 660

Qy        642 IPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRK 701
              || |||||||:: | :|||| ||||||:| ||: ||||:||:|||||:|:|||||||||
Db        661 IPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRINTQFNKRK 720

Qy        702 EQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRF 761
              ||||||||||:|:|  || |||:|| :||| | :|:|  || |||:|||:|| ||:||
Db        721 EQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESKYNINNIEEELNKKVSLAMKNIERF 780

Qy        762 LTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNN 821
              :|||||||||||||| :: ||:|||::||: |||||: |  ||||  :||  :|  |||:
Db        781 MTESSISYLMKLINEAEVGKLKEYDKHVKSDLLDYILYHKLILGEQTKELIDLVTSTLNS 840

Qy        822 SIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKY 881
              ||||:|||||:||||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| |
Db        841 SIPFELSSYTNDKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIY 900

Qy        882 PTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIIN 941
              ||:|||||: :||||||:||: |||:::|:|||||||| || :  : :| | ||||||
Db        901 STNRNQFGIYSGRLSEVNIAQNNDIIYNSRYQNFSISFWVTIPKH-YRPMNRNREYTIIN 959

Qy        942 CMRDNNSGWKVSL---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITND 998
              || :||||||:||       |||||||| ||  :|| | |     |||||||||||||:
Db        960 CMGNNNSGWKISLRTIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNN 1019

Qy        999 RLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETE 1058
              |||:|::||||||| :||| |||:||||||||||| |    |:|||| :|: |||:||
Db       1020 RLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTE 1079

Qy       1059 IQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRS----- 1113
              |:|||:|||: :||||:|||||||:|:||| |:|: : :| | :    |:|| |
Db       1080 IETLYSNEPDPSILKDYWGNYLLYNKKYYLFNLLRKDKYITRNS-GILNINQQRGVTGGI 1138

Qy       1114 TILLANRLYSGIKVKIQR---VNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTN 1170
              :: |  :|| |::| |::   :: |:| || |||||  ||| |         |||| : |
Db       1139 SVFLNYKLYEGVEVIIRKNAPIDISNT-DNFVRKNDLAYINVVDHGVE-YRLYADISITK 1196

Qy       1171 KEKTIKI---SSSGNRFNQVVVMNSVG--CTMNFKNNNGNNIGLLGFKADTVVASTWYYT 1225
              || ||:   |:  :   |::||:|:|  |||||:||:|:|||||| :| :|||:|||
Db       1197 SEKIIKLIRTSNPNDSLGQIIVMDSIGNNCTMNFQNNDGSNIGLLGFHSDDLVASSWYYN 1256

Qy       1226 HMRDNTNSNGFFWNFISEEHGWQE 1249
              |:| ||:||| ||:|||:||||:|
Db       1257 HIRRNTSSNGCFWSFISKEHGWKE 1280


RESULT 7
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
```

```
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             34.2%;  Score 2228;  DB 2;  Length 1291;
  Best Local Similarity   38.2%;  Pred. No. 4.5e-78;
  Matches  501;  Conservative  265;  Mismatches  461;  Indels   86;  Gaps   33;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || |||:|||||::|  |: ::|    |  ::||:| |   |||||||   | |:||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
              : : | |   ||||:|| :: :|: ||: : |:||||     |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   |  :|  | ||   :  :    :    :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   | :   |:          : |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :      ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |     |  :| ||: |: ||   :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
              ::|  :::| | :|       |: |||:: || | ||:||:  |::  :||||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDD-NINTPK 452
              :    |:   |    |: ||:: :       |||:::| :|||:| :|||::|| : |
Db        419 QAYEEISKEHLAVYKIQMCKSVRA-----PGICIDVDNEDLFFIADKNSFSDDLSKNERI 473

Qy        453 EIDDTVTSNNNYEN--DLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIE 510
              | |  | :|  ||   :|::||: |  |   |  |         :  :| |:      |:
Db        474 EYD---TQSNYIENRSSIDELILDTNLISKIELPSENTESLTDFNVDVPVYEKQPA--IK 528

Qy        511 QHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAA 570
              :  :|  :| || :|  |    :::|||| | |||   |:|:||| ::|   || |:|
Db        529 KIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAG 588

Qy        571 LFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGA 630
              ||  |::|:: ||  |||: ||:|||||||::|||||||||:|||  ||||::| |: ||
Db        589 LFAGWVKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGA 648

Qy        631 GILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWM 690
              ||||| ||||| :  | ::|::   |||||:|| |:||| :|||| ::|  ||: |:
Db        649 SILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRDEKWIDMYGLIVAQWL 705

Qy        691 TKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQK 750
              : :||||   || ||:|| |  ||: ||: ||| |:.:||: :    |   | ::||:
Db        706 STVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNIYSEKEKSNI--NIDFNDINSKLNEG 763

Qy        751 VSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQE 810
              :: |::||: |: | |:||||| :  : : || ::|  :|  ||:|| ::   |  | :
Db        764 INQAVDNINNFINECSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEY 823

Qy        811 LNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDS 870
              | |   |  ||| || ||:: |||  |||:   | :: :||:||:::  :| ||| :
Db        824 EKSKVDKHLKTIIPFDLSMYTNNTILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGA 883

Qy        871 NININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKI 930
              |: :   |    :|||| : :   ||: ::||  ||::: : :||:|||:||| | |
Db        884 NVEVYDGVEL--NDKNQFKLTSSTNSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDG 941

Qy        931 VN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYIN 988
              :   ::|||||||:: ||||||:|:  | ||||| | :|  : : | |      ||||||
Db        942 IQNYIHNEYTIINCIK-NNSGWKISIRGNRIIWTLTDINGKTKSVFFEYSIREDISDYIN 1000

Qy        989 KWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYF 1047
              :| ||||||:   ::|:|||| |     | ::| :  :  |:||:  :   |::| ::||
Db       1001 RWFFVTITNNS-DNAKIYINGKLESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYF 1059

Qy       1048 NIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS--- 1104
              :||: || :: |: :|  :  :  ||||||| |:|:|||| |    |::|  : ||
Db       1060 SIFNTELSQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVG 1119

Qy       1105 ---TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLL 1160
                 | |  |  |  :     || | |  |:| :|| | ||::||| | :|::|  |
Db       1120 EILTRSKYNQNSNYINYRNLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNSNRE-W 1178

Qy       1161 PLYADTATTNKEKTIKISS--SGNRFNQVVVMN------SVGCTMNFKNN--NGNNIGLL 1210
                :||      :|| : :::    | | : : :       :  | : || :  : : |||:
Db       1179 RVYAYKDFKEEEKKLFLANIYDSNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLI 1238

Qy       1211 G----FKADTVV--------ASTWYYTHM-RDNTNSN-GFFWNFISEEHGWQE 1249
              |    :::  |:       | ||   :  |  | | |  | || :: || |
Db       1239 GIHRFYESGIVLKDYKNYFCISKWYLKEVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 8
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             33.9%;  Score 2207;  DB 2;  Length 1291;
  Best Local Similarity   38.1%;  Pred. No. 2.9e-77;
  Matches  500;  Conservative  261;  Mismatches  468;  Indels   82;  Gaps   33;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
              || |||:|||||::|  |: ::|    |  ::||:| |   |||||||   | |:||
```

```
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
Db        240 YGIK-VNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
Db        473 IAYN-TQNNYIENDFSINELILDTDLISKIELPSENTESLTDFNVYVPVYKKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
Db        590 FAGWVKQIVNDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAPEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIETINSALTKRDEKWIDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIENELNQKV 751
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNIYSEKERSNI--NIDFNDVNSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILGESQQEL 811
Db        765 NQAIDNINNFINECSVSYLMKKMIPLAVEKLLDFDNTLRKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
Db        825 KSKVDKYLKTSIPFDLSTYTNNTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
Db        885 VEVYDGVKL--NDKNQFKLTSSANSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINK 989
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
Db       1002 WFFVTITNNS-DNAKIYINGKLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDS---- 1104
Db       1061 IFNTELSQSNIEEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLLP 1161
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRV 1180

Qy       1162 LYADTATTNKEKTI--KISSSGNRFNQVVVMN-----SVGCTMNFKNN--NGNNIGLLG- 1211
Db       1181 YMYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGI 1240

Qy       1212 ---------FK--ADTVVASTWYYTHM-RDNTNSN-GFFWNFISEEHGWQE 1249
Db       1241 HRFYESGIVFKEYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291
```

RESULT 9
Q9ZAJ8

```
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;
```

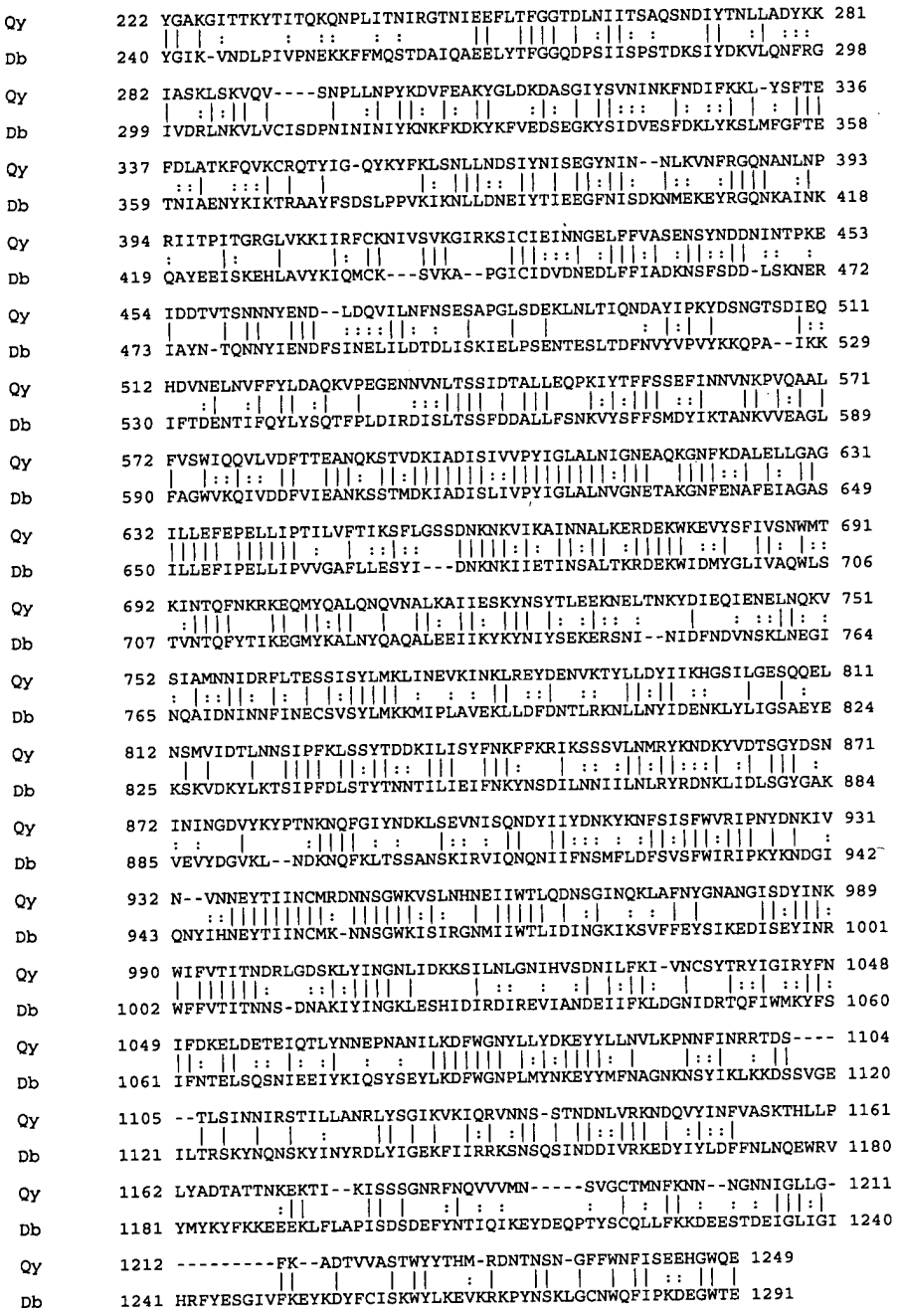
```
  Query Match           33.8%;  Score 2203;  DB 2;  Length 1291;
  Best Local Similarity  38.1%;  Pred. No. 4.2e-77;
  Matches  500;  Conservative  259;  Mismatches  470;  Indels   82;  Gaps   33;

Qy          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

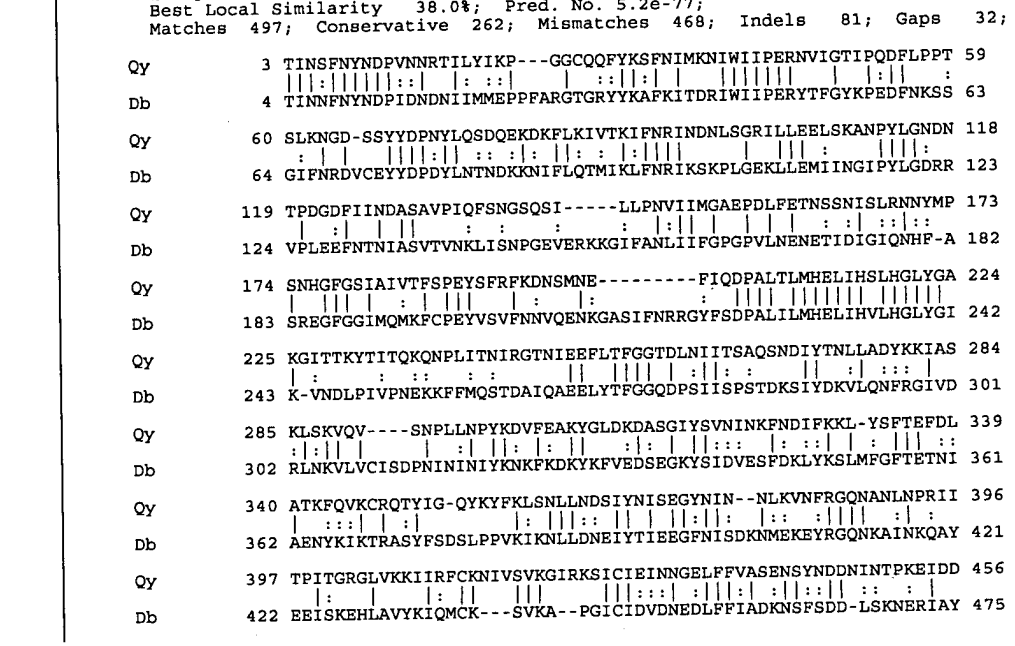
Qy        116 NDNTPDGDFIINDASAVPIQFSNGSQSI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

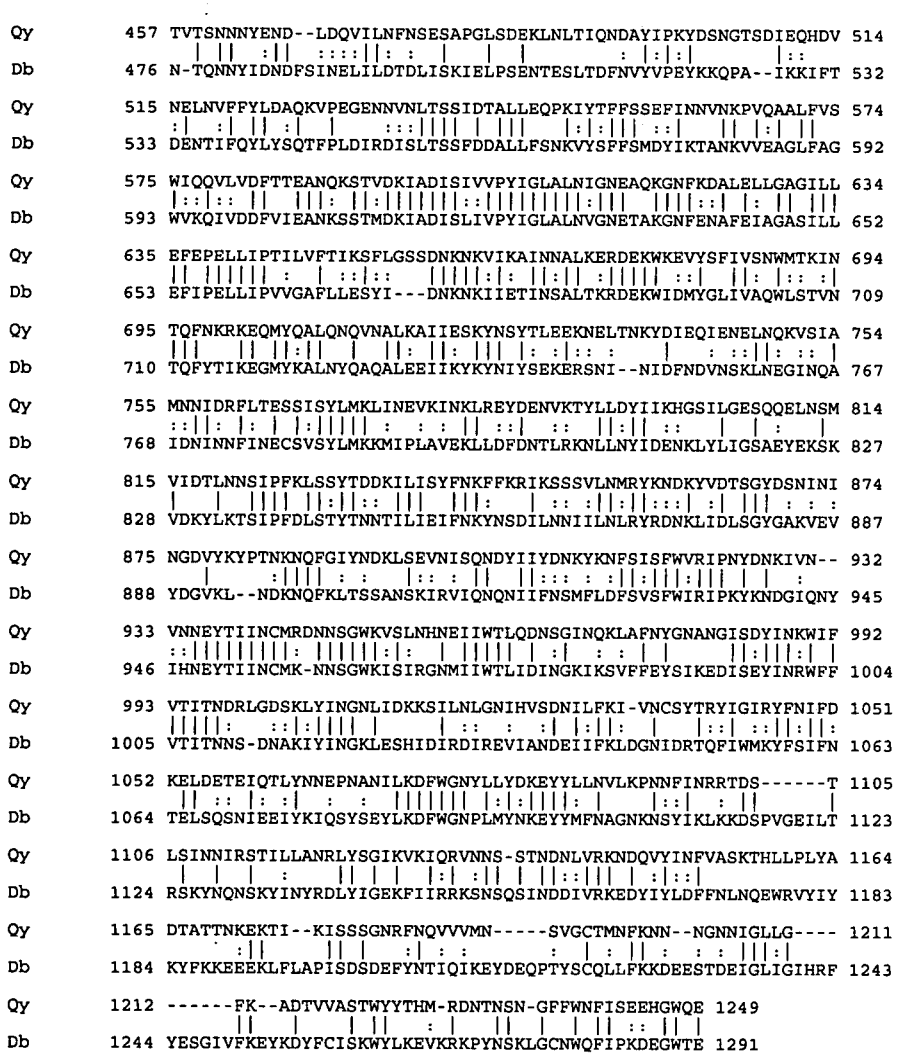
Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHGL 221
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239
```



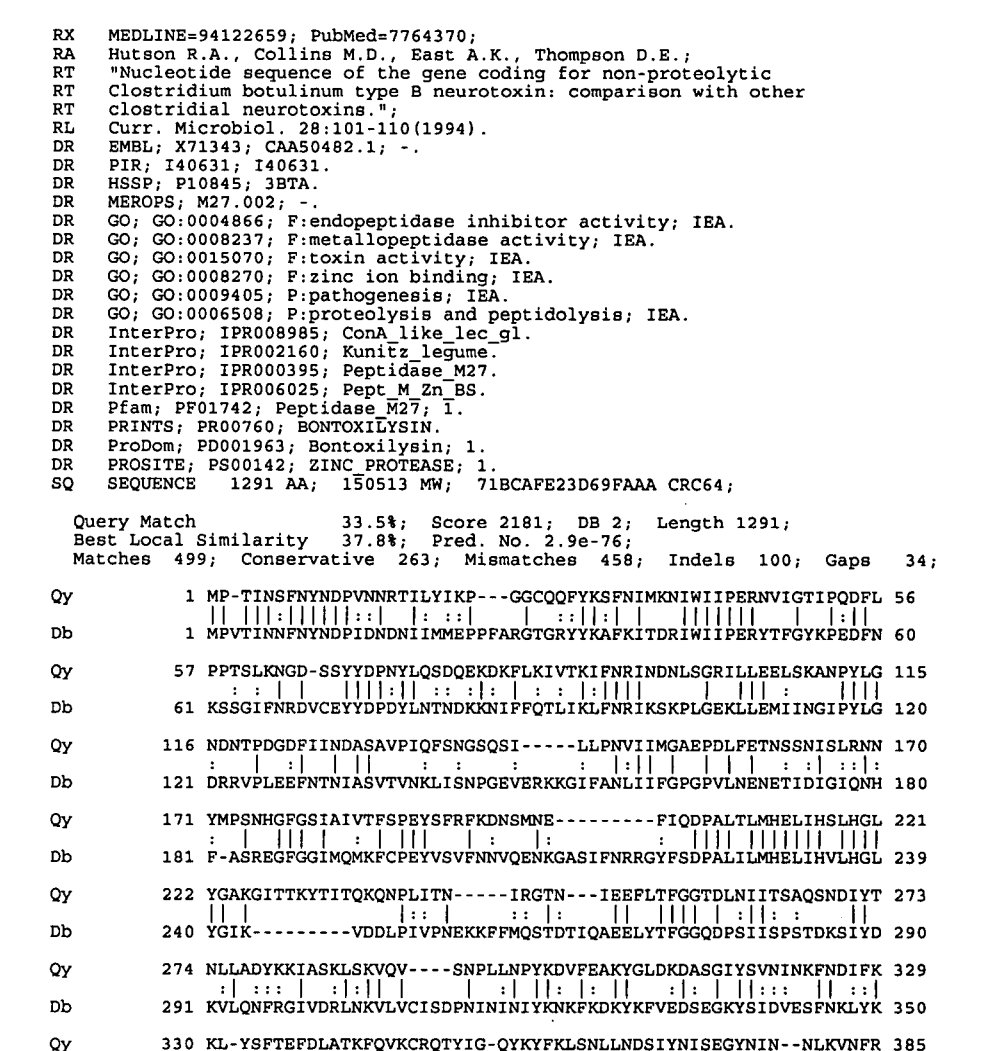
```
RESULT 10
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA. .
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match           33.8%;  Score 2200.5;  DB 2;  Length 1291;
  Best Local Similarity  38.0%;  Pred. No. 5.2e-77;
  Matches  497;  Conservative  262;  Mismatches  468;  Indels   81;  Gaps   32;
```



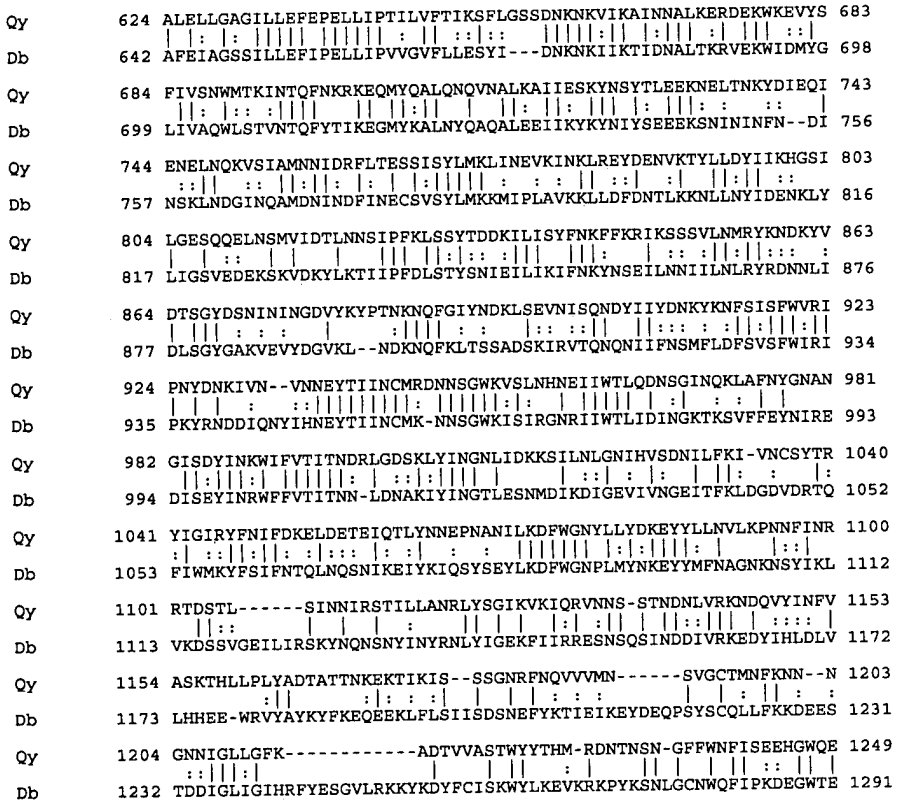


RESULT 11
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;



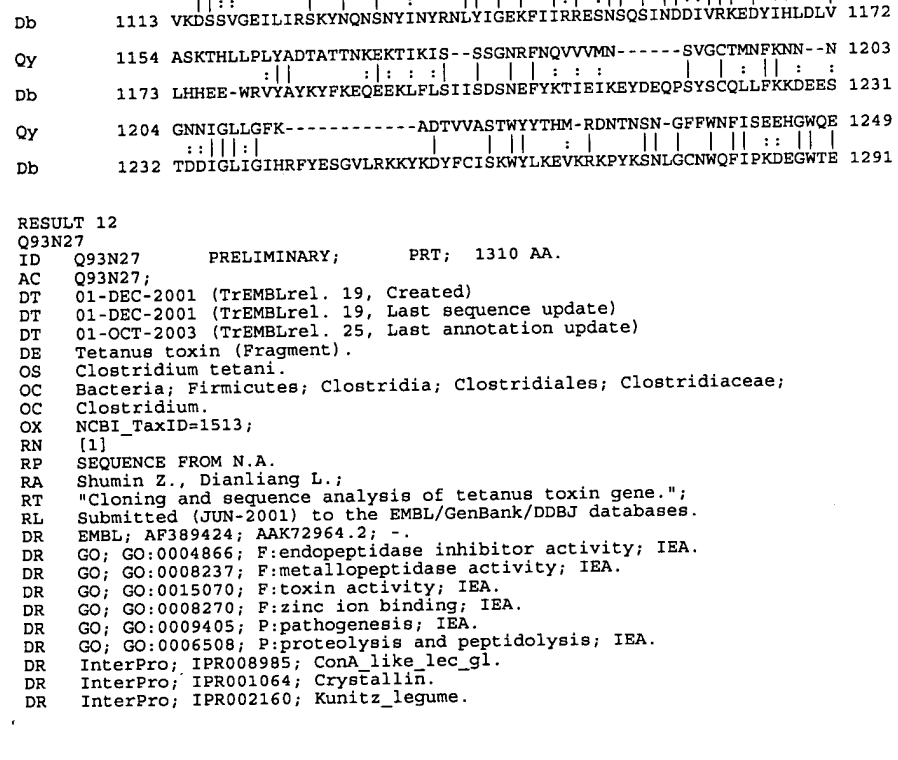
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR PIR; I40631; I40631.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 33.5%; Score 2181; DB 2; Length 1291;
Best Local Similarity 37.8%; Pred. No. 2.9e-76;
Matches 499; Conservative 263; Mismatches 458; Indels 100; Gaps 34;



RESULT 12
Q93N27
ID Q93N27 PRELIMINARY; PRT; 1310 AA.
AC Q93N27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tetanus toxin (Fragment).
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RA Shumin Z., Dianliang L.;
RT "Cloning and sequence analysis of tetanus toxin gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF389424; AAK72964.2; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1 1
FT NON_TER 1310 1310
SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 32.4%; Score 2111; DB 2; Length 1310;
Best Local Similarity 35.7%; Pred. No. 1.5e-73;
Matches 488; Conservative 254; Mismatches 432; Indels 194; Gaps 38;



```
Db        828 ILMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL----- 882

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSEVNI 900

Db        883 ------KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIV 936

Qy        901 SQNDYIIYDNKYKNFSISFWVRIP--------NYDNKIVNVNNEYTIINCMRDNN----S 948

Db        937 HKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYD------TNEYSIISSMKKYSLSIGS 990

Qy        949 GWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISD----YI-NKWIFVTITNDRLGDS 1003

Db        991 GWSVSLKGNNLIWTLKDSAGEVRQITF-----RDLSDKFNAYLANKWVFITITNDRLSSA 1045

Qy       1004 KLYINGNLIDKKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTL 1062

Db       1046 NLYINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKL 1105

Qy       1063 YNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLAN--- 1119

Db       1106 YTSYLSITFLRDFWGNPLRYDTEYYLIPV--------AYSSKDVQLKNITDYMYLTNAPS 1157

Qy       1120 ----------RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLLPLYADT 1166

Db       1158 YTNGKLNIYYRRLYSGLKFIIKRYTPNNEIDSFVRSGDFIKLYVSY-------------- 1203

Qy       1167 ATTNKEKTIKISSSGNRFNQV-----VVMNSVGCTMNFK------------------NN 1202

Db       1204 --NNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDD 1261

Qy       1203 NGNNIGLLGF---------KADTVVASTWYYTHMRDNTNSNGFFWNFI 1241

Db       1262 KDASLGLVGTHNGQIGNDPNRDILIASNWYFNHLKDKTLTCD--WYFV 1307


RESULT 13
Q45862
ID   Q45862      PRELIMINARY;      PRT;   367 AA.
AC   Q45862;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type E;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70815; CAA50146.1; -.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42854 MW;  0810595B3A865570 CRC64;

  Query Match            28.8%;  Score 1875;  DB 2;  Length 367;
  Best Local Similarity   98.1%;  Pred. No. 4.4e-65;
  Matches  360;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELT 735

Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLD 795

Db        121 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855

Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915

Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAF 975

Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982

Db        361 NYGNANG 367


RESULT 14
Q45861
ID   Q45861      PRELIMINARY;      PRT;   367 AA.
AC   Q45861;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type E;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70818; CAA50149.1; -.
DR   PIR; S48106; S48106.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42902 MW;  346A610C2FF70262 CRC64;

  Query Match            28.7%;  Score 1866;  DB 2;  Length 367;
  Best Local Similarity   97.8%;  Pred. No. 9.7e-65;
  Matches  359;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELT 735

Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLD 795

Db        121 NKYDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180
```

```
Qy        796 YIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
              |||:||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAF 975
              |||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
              |||||||
Db        361 NYGNANG 367
```

```
RESULT 15
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match           28.6%;  Score 1860;  DB 12;  Length 1275;
  Best Local Similarity   33.6%;  Pred. No. 6.6e-64;
  Matches  444;  Conservative  259;  Mismatches  491;  Indels  126;  Gaps   35;

Qy          4 INSFNYNDPVNNRTILYIK---------PGGCQQFYKSFNIMKNIWIIPERNVIGTIPQD 54
              :  |||:||||:  |||::          |       |:| | :|||:||||    | |
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FLPPTSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYL 114
                ||        |||||:|| :|::|| ||| : |:| |||:   |: |:  |   :|::
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDG--DFIINDASAVPIQFSNGS---QSILLPNVIIMGAEPDLFETNSSNISLRN 169
              |: :||:   ||  :  :    :| |||    :|: |:|:| |  |:: :  :|
Db        118 GDSSTPEDTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQ 177

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFKDNSMNE---------FIQDPALTLMHELIHSLHG 220
                ||  |||:::|:  :||:   | | : |:          |  || : ||||| ||||
Db        178 QSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQ 237

Qy        221 LYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
              |||   | :   |  : :    :  | |:  ||  |||| |: ||   : : :    |
Db        238 LYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGH 296

Qy        279 YKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS 333
              || || :|: :  : |       :: || :|  ||  ||| :| : |||:||| ::  | :
Db        297 YKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTN 356

Qy        334 -FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNAN 390
               :|   :::: || |  |  ::     :|:|:|:|| | :|:|: |    |   |||
Db        357 VMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIE 416

Qy        391 LNPRIITPITGRGLVKKIIRFCKNIVSVKGIR-KSICIEINNGELFFVASENSYNDDNIN 449
              ||  :  ::   :|    : | :   |  | | ||:: |  | :|| ::| : :
Db        417 RNP-ALQKLSSESVVDLFTKVCLRL--TKNSRDDSTCIKVKNNRLPYVADKDSISQEIFE 473

Qy        450 TPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPKY 501
                 |:|   | : :  ||: ||:     | |       | :: | ||| :   :   |
Db        474 NKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVF---Y 530

Qy        502 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 561
              |      || :: |: || ::||::||:     |: ||:|:: ||    |||||  |
Db        531 D-----DITKY-VDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPS-LAE 583

Qy        562 NVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 621
              ||| ||| ||::|  :|: ||||  .:|  |:|||:|:|:::|||| |||||| | :|||
Db        584 KVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNF 643

Qy        622 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 681
              |    |  |||  ||  || : |||    |  |   : |:|| | | |::| ::||:
Db        644 NQAFATAGVAFLLEGFPEFTIPALGVFT---FYSSIQEREKIIKTIENCLEQRVKRWKDS 700

Qy        682 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIE 741
              | ::|||:::| ||||    ||| :|  | :|:|| |: :|  |:  :|  :  |  :|
Db        701 YQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENI--KSQVE 758

Qy        742 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHG 801
              ::| |: |:| |||||::|: | |::|| | :    |::| ::|   || |:: |  |
Db        759 NLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHN 818

Qy        802 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 861
              ||      | : | ::  |::|| : |||:: :|    |::|  |  | :|::: | :
Db        819 IILVGEVDRLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNA 878

Qy        862 YVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS----EVNISQNDYIIYDNKYKNFSI 917
              ||||||:: : :  :|      | |     |||    :: :: |: |:|   |:| |:
Db        879 LVDTSGYNAEVRVGDNVQLNTIYTNDF-----KLSSSGDKIIVNLNNNILYSAIYENSSV 933

Qy        918 SFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNY 977
              |||::|      : | :|||||| : : |||||: : :  | | ||| :   : | |:|
Db        934 SFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDY 989

Qy        978 GNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VNC 1036
                :  : | ||| ||||||: :|  ||||||  |   : | :|  : :   |:| |  |
Db        990 SESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENI 1049

Qy       1037 SYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNN 1096
              \  : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|    :|
Db       1050 DENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DN 1105

Qy       1097 FINRRTDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYI 1150
              :|:|              | :|:       ::||:|  :  |: |:: :    ::  :: :
Db       1106 YIDRYIAP-------ESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILH 1158

Qy       1151 NFVASKTHLLPLYADT--------ATTNKEKTIKISSSGNRFN------QVVVMNSVGCT 1196
                |: :::    ||         :|  :|: |:   :    :  :|  :  |:
Db       1159 MLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCS 1218

Qy       1197 MNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGW 1247
```

```
                 | :   | : |          || | | || | | | :     |   || ||| : ||
Db       1219 QIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFISRDPGW 1274
```

Search completed: September 10, 2004, 03:41:33
Job time : 62.1937 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 73.3569 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 1252
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	804	64.2	804	5	ABG69083	Abg69083 Botulinum
2	422	33.7	422	5	ABG69071	Abg69071 Botulinum
3	381	30.4	382	4	AAB04100	Aab04100 Botulism
4	361	28.8	452	2	AAW68396	Aaw68396 Clostridi
5	354	28.3	419	4	AAB04095	Aab04095 Botulism
6	354	28.3	449	3	AAY77137	Aay77137 Synthetic
7	354	28.3	449	4	AAB04094	Aab04094 Botulism
8	331	26.4	427	5	ABG69082	Abg69082 Botulinum
9	140	11.2	451	2	AAW68395	Aaw68395 Clostridi
10	20	1.6	408	4	AAB04101	Aab04101 Botulism
11	20	1.6	858	5	ABG69085	Abg69085 Botulinum
12	20	1.6	1059	3	AAY93309	Aay93309 A mangane
13	20	1.6	1084	3	AAY93312	Aay93312 A mangane
14	15	1.2	144	2	AAW09016	Aaw09016 Immunogen
15	15	1.2	431	2	AAW09014	Aaw09014 Immunogen
16	15	1.2	432	3	AAY77138	Aay77138 Synthetic
17	15	1.2	432	4	AAB04103	Aab04103 Botulism
18	15	1.2	432	4	AAB04096	Aab04096 Botulism
19	15	1.2	645	4	AAE07894	Aae07894 Modified
20	15	1.2	645	6	AAE35692	Aae35692 DipT HN d
21	15	1.2	657	6	AAE35693	Aae35693 BoNT/F-Hc
22	15	1.2	657	6	AAE35694	Aae35694 BoNT/F-Hc
23	15	1.2	660	4	AAE07898	Aae07898 Modified
24	15	1.2	685	4	AAE07893	Aae07893 Modified
25	15	1.2	862	4	AAE07890	Aae07890 Modified
26	15	1.2	887	4	AAE07892	Aae07892 Modified
27	15	1.2	979	6	AAE35713	Aae35713 BoNT/F-Hc
28	15	1.2	1032	4	AAE07901	Aae07901 C. botuli
29	15	1.2	1092	4	AAE07900	Aae07900 C. botuli
30	15	1.2	1192	6	AAE35711	Aae35711 BoNT/F-Hc
31	15	1.2	1192	6	AAE35710	Aae35710 BoNT/F-Hc
32	12	1.0	436	5	ABG69072	Abg69072 Botulinum
33	12	1.0	439	5	ABG69073	Abg69073 Botulinum
34	12	1.0	443	5	ABG69084	Abg69084 Botulinum
35	12	1.0	444	5	ABG69086	Abg69086 Botulinum
36	12	1.0	448	2	AAW68399	Aaw68399 Clostridi
37	12	1.0	848	5	ABG69087	Abg69087 Botulinum
38	11	0.9	144	2	AAW09015	Aaw09015 Immunogen
39	10	0.8	233	3	AAY77143	Aay77143 Botulinum
40	10	0.8	382	3	AAB36303	Aab36303 BoNT/A pr
41	10	0.8	415	4	AAB04083	Aab04083 Botulism
42	10	0.8	432	3	AAY77142	Aay77142 Native bo
43	10	0.8	434	4	AAB04089	Aab04089 Botulism
44	10	0.8	435	4	AAB04090	Aab04090 Botulism
45	10	0.8	437	4	AAB04088	Aab04088 Botulism

ALIGNMENTS

RESULT 1
ABG69083
ID ABG69083 standard; protein; 804 AA.
XX
AC ABG69083;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #17.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98553.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 152-154; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

```
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 804 AA;

  Query Match             64.2%;  Score 804;  DB 5;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  804;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSIL 804
              ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSIL 804


RESULT 2
ABG69071
ID   ABG69071 standard; protein; 422 AA.
XX
AC   ABG69071;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #5.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98541.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 125-126; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 422 AA;

  Query Match             33.7%;  Score 422;  DB 5;  Length 422;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IR 422
                 ||
Db        421 IR 422
```

RESULT 3
AAB04100
ID AAB04100 standard; protein; 382 AA.
XX
AC AAB04100;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain N-terminal sequence (serotype E).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54496.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 6; Fig 15b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 382 AA;

```
  Query Match             30.4%;  Score 381;  DB 4;  Length 382;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        424 SICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGL 483
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGL 61

Qy        484 SDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDT 543
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 SDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDT 121

Qy        544 ALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVV 603
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVV 181

Qy        604 PYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKV 663
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 PYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKV 241

Qy        664 IKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKY 723
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 IKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKY 301

Qy        724 NSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLR 783
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 NSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLR 361

Qy        784 EYDENVKTYLLNYIIQHGSIL 804
                 |||||||||||||||||||
Db        362 EYDENVKTYLLNYIIQHGSIL 382
```

RESULT 4
AAW68396
ID AAW68396 standard; protein; 452 AA.
XX
AC AAW68396;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain NCTC 11219.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .21
FT   /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;

```
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;

  Query Match             28.8%;  Score 361;  DB 2;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  361;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        892 NDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWK 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         92 NDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWK 151

Qy        952 VSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        152 VSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 211

Qy       1012 IDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNI 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 IDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNI 271

Qy       1072 LKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQR 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 LKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQR 331

Qy       1132 VNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMN 1191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 VNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMN 391

Qy       1192 SVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 SVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 451

Qy       1252 K 1252
              |
Db        452 K 452


RESULT 5
AAB04095
ID   AAB04095 standard; protein; 419 AA.
XX
AC   AAB04095;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54489.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Disclosure; Fig 8; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 419 AA;

  Query Match             28.3%;  Score 354;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  354;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        899 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 125

Qy        959 IIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 IIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL 185

Qy       1019 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGN 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGN 245

Qy       1079 YLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 YLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 305

Qy       1139 DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCT 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCT 365

Qy       1199 MNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 MNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 419


RESULT 6
AAY77137
ID   AAY77137 standard; protein; 449 AA.
XX
AC   AAY77137;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype E (BoNTE) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype E; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87215.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 26; Page 43-44; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 449 AA;

  Query Match             28.3%;  Score 354;  DB 3;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  354;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        899 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 155

Qy        959 IIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 IIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL 215

Qy       1019 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGN 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGN 275

Qy       1079 YLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 YLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 335

Qy       1139 DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCT 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCT 395

Qy       1199 MNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 MNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 449


RESULT 7
AAB04094
ID   AAB04094 standard; protein; 449 AA.
XX
AC   AAB04094;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54488.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 7b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
```

```
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 449 AA;

  Query Match             28.3%;  Score 354;  DB 4;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  354;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        899 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 NISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNE 155

Qy        959 IIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 IIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL 215

Qy       1019 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGN 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGN 275

Qy       1079 YLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 YLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTN 335

Qy       1139 DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCT 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCT 395

Qy       1199 MNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 MNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 449


RESULT 8
ABG69082
ID   ABG69082 standard; protein; 427 AA.
XX
AC   ABG69082;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #16.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98552.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 150-151; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 427 AA;

  Query Match             26.4%;  Score 331;  DB 5;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         93 RINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMG 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 RINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMG 152

Qy        153 AEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNEFIQDPALTLMH 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 AEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNEFIQDPALTLMH 212

Qy        213 ELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIY 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 ELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIY 272

Qy        273 TNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLY 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 TNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLY 332

Qy        333 SFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLN 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 SFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLN 392

Qy        393 PRIITPITGRGLVKKIIRFCKNIVSVKGIRK 423
              |||||||||||||||||||||||||||||||
Db        393 PRIITPITGRGLVKKIIRFCKNIVSVKGIRK 423


RESULT 9
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
```

```
FT                  /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             11.2%;  Score 140;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-139;
  Matches  140;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        827 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 86

Qy        887 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 146

Qy        947 NSGWKVSLNHNEIIWTLQDN 966
              ||||||||||||||||||||
Db        147 NSGWKVSLNHNEIIWTLQDN 166


RESULT 10
AAB04101
ID   AAB04101 standard; protein; 408 AA.
XX
AC   AAB04101;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain N-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54497.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 6; Fig 16b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 408 AA;

  Query Match             1.6%;  Score 20;  DB 4;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 1e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        273 INTQFNKRKEQMYQALQNQV 292


RESULT 11
ABG69085
ID   ABG69085 standard; protein; 858 AA.
XX
AC   ABG69085;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #19.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
```

PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98555.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 157-159; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 858 AA;

Query Match 1.6%; Score 20; DB 5; Length 858;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 723 INTQFNKRKEQMYQALQNQV 742

RESULT 12
AAY93309
ID AAY93309 standard; protein; 1059 AA.
XX
AC AAY93309;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 48-51; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype F. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a domain
CC that translocates the SOD of the composition into the neuronal cell.
CC After translocation, the linker is cleaved to release the SOD. The
CC composition is useful for treating neuronal diseases caused or augmented
CC by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC disease, Huntington's disease and motor neurone diseases
XX
SQ Sequence 1059 AA;

Query Match 1.6%; Score 20; DB 3; Length 1059;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 INTQFNKRKEQMYQALQNQV 712
||||||||||||||||||||
Db 493 INTQFNKRKEQMYQALQNQV 512

RESULT 13
AAY93312
ID AAY93312 standard; protein; 1084 AA.
XX
AC AAY93312;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.

```
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 57-60; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide dismutase
CC   (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC   cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC   serotype F. The specification describes a composition for delivery of SOD
CC   to neuronal cells. The composition comprises SOD linked, by a cleavable
CC   linker, to a neuronal cell targeting component (NCTC). This component has
CC   a domain that binds to a neuronal cell and a domain that translocates the
CC   SOD of the composition into the neuronal cell. After translocation, the
CC   linker is cleaved to release the SOD. The composition is useful for
CC   treating neuronal diseases caused or augmented by oxidative stress, such
CC   as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases
XX
SQ   Sequence 1084 AA;

  Query Match             1.6%;  Score 20;  DB 3;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        518 INTQFNKRKEQMYQALQNQV 537


RESULT 14
AAW09016
ID   AAW09016 standard; protein; 144 AA.
XX
AC   AAW09016;
XX
DT   17-OCT-2003  (revised)
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin polypeptide (aa992-1135).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum; type F strain Langeland.
XX
PN   WO9641881-A1.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-GB001409.
XX
PR   12-JUN-1995;   95GB-00011909.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
XX
DR   WPI; 1997-065467/06.
XX
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT   vaccine prodn.
XX
PS   Claim 5; Page 18-19; 37pp; English.
XX
CC   Novel polypeptides (AAW09014-17) respectively comprise amino acids 848-
CC   1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a type F
CC   botulinum neurotoxin (BoNT/F). They lack the L chain and HN epitopes
CC   necessary for metalloprotease activity and toxin internalisation. They
CC   are free of botulinum toxin activity but can induce protective immunity
CC   to a type F botulinum toxin, making them useful for vaccine prodn.
CC   Recombinant polypeptides can be produced in transformed host cells, esp.
CC   as fusion proteins, e.g. with maltose binding protein to facilitate
CC   purification. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 144 AA;

  Query Match             1.2%;  Score 15;  DB 2;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        983 ISDYINKWIFVTITN 997
              |||||||||||||||
Db         10 ISDYINKWIFVTITN 24


RESULT 15
AAW09014
ID   AAW09014 standard; protein; 431 AA.
XX
AC   AAW09014;
XX
DT   17-OCT-2003  (revised)
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum; type F strain Langeland.
XX
PN   WO9641881-A1.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-GB001409.
XX
PR   12-JUN-1995;   95GB-00011909.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
XX
DR   WPI; 1997-065467/06.
DR   N-PSDB; AAT48100.
XX
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT   vaccine prodn.
XX
PS   Claim 5; Page 16-17; 37pp; English.
XX
CC   A polypeptide (AAW09014) comprises the heavy chain (amino acids 848-1278)
CC   of a type F botulinum neurotoxin (BoNT/F), and can be produced using a
CC   synthetic gene (AAT48101) based on the natural gene sequence (AAT48100)
CC   for the heavy chain. The polypeptides and its fragments (see also
CC   AAW09015-17) lack the light chain and HN epitopes necessary for
CC   metalloprotease activity and toxin internalisation. They are free of
CC   botulinum toxin activity but can induce protective immunity to a type F
CC   botulinum toxin, making them useful for vaccine prodn. Recombinant
CC   polypeptides can be produced in transformed host cells, esp. as fusion
CC   proteins, e.g. with maltose binding protein to facilitate purification.
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 431 AA;

  Query Match             1.2%;  Score 15;  DB 2;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        983 ISDYINKWIFVTITN 997
              |||||||||||||||
Db        154 ISDYINKWIFVTITN 168


Search completed: September 10, 2004, 05:05:57
```

Job time : 74.3569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 20.5272 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 1252
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	0.8	438	1	US-08-480-604A-23	Sequence 23, Appl
2	10	0.8	438	2	US-08-405-496A-23	Sequence 23, Appl
3	10	0.8	438	3	US-08-915-136-23	Sequence 23, Appl
4	10	0.8	438	4	US-09-084-517-23	Sequence 23, Appl
5	10	0.8	462	1	US-08-480-604A-26	Sequence 26, Appl
6	10	0.8	462	2	US-08-405-496A-26	Sequence 26, Appl
7	10	0.8	462	3	US-08-915-136-26	Sequence 26, Appl
8	10	0.8	462	4	US-09-084-517-26	Sequence 26, Appl
9	10	0.8	1296	1	US-08-480-604A-28	Sequence 28, Appl
10	10	0.8	1296	2	US-08-405-496A-28	Sequence 28, Appl
11	10	0.8	1296	3	US-08-915-136-28	Sequence 28, Appl
12	10	0.8	1296	4	US-09-084-517-28	Sequence 28, Appl
13	9	0.7	858	4	US-09-255-829-22	Sequence 22, Appl
14	9	0.7	858	4	US-09-255-829-29	Sequence 29, Appl
15	9	0.7	1169	4	US-09-255-829-20	Sequence 20, Appl
16	8	0.6	260	2	US-07-857-224B-27	Sequence 27, Appl
17	8	0.6	466	4	US-09-134-001C-3355	Sequence 3355, Ap
18	8	0.6	473	4	US-09-252-991A-28845	Sequence 28845, A
19	8	0.6	793	4	US-09-107-532A-6223	Sequence 6223, Ap
20	7	0.6	17	1	US-08-149-809-2	Sequence 2, Appli
21	7	0.6	35	4	US-09-266-764-7	Sequence 7, Appli
22	7	0.6	68	4	US-09-134-000C-4208	Sequence 4208, Ap
23	7	0.6	79	2	US-09-196-048-2	Sequence 2, Appli
24	7	0.6	79	4	US-09-134-000C-6016	Sequence 6016, Ap
25	7	0.6	177	4	US-09-252-991A-29795	Sequence 29795, A
26	7	0.6	199	4	US-09-134-001C-3825	Sequence 3825, Ap
27	7	0.6	199	4	US-09-134-001C-4609	Sequence 4609, Ap
28	7	0.6	260	2	US-07-857-224B-28	Sequence 28, Appl
29	7	0.6	283	4	US-09-198-452A-803	Sequence 803, App
30	7	0.6	335	4	US-09-489-039A-11298	Sequence 11298, A
31	7	0.6	359	4	US-09-403-343B-21	Sequence 21, Appl
32	7	0.6	411	4	US-09-328-352-6719	Sequence 6719, Ap
33	7	0.6	417	4	US-09-252-991A-16876	Sequence 16876, A
34	7	0.6	431	4	US-09-372-668-4	Sequence 4, Appli
35	7	0.6	431	4	US-09-697-377-4	Sequence 4, Appli
36	7	0.6	431	4	US-09-696-868-4	Sequence 4, Appli
37	7	0.6	431	4	US-09-697-341-4	Sequence 4, Appli
38	7	0.6	434	4	US-09-252-991A-23499	Sequence 23499, A
39	7	0.6	438	4	US-09-134-000C-4760	Sequence 4760, Ap
40	7	0.6	452	1	US-07-618-312A-2	Sequence 2, Appli
41	7	0.6	452	1	US-07-618-312A-4	Sequence 4, Appli
42	7	0.6	452	1	US-08-110-786A-8	Sequence 8, Appli
43	7	0.6	452	1	US-08-280-228-2	Sequence 2, Appli
44	7	0.6	452	1	US-08-280-228-4	Sequence 4, Appli
45	7	0.6	502	4	US-09-796-202-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER:• IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match             0.8%;  Score 10;  DB 1;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 2
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match             0.8%;  Score 10;  DB 2;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 3
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match             0.8%;  Score 10;  DB 3;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
```

```
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 4
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-23

  Query Match             0.8%;  Score 10;  DB 4;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        101 NNEYTIINCM 110


RESULT 5
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             0.8%;  Score 10;  DB 1;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        125 NNEYTIINCM 134


RESULT 6
US-08-405-496A-26
; Sequence 26, Application US/08405496A
```

; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-26

Query Match 0.8%; Score 10; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 934 NNEYTIINCM 943
||||||||||
Db 125 NNEYTIINCM 134

RESULT 7
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-26

Query Match 0.8%; Score 10; DB 3; Length 462;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 934 NNEYTIINCM 943
||||||||||
Db 125 NNEYTIINCM 134

RESULT 8
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104

```
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,517
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CARROLL, PETER G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  OPHD-01610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  462 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             0.8%;  Score 10;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        125 NNEYTIINCM 134


RESULT 9
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;     APPLICANT:  KINK, JOHN A.
;     APPLICANT:  THALLEY, BRUCE S.
;     APPLICANT:  PADHYE, NISHA V.
;     APPLICANT:  FIRCA, JOSEPH R.
;     APPLICANT:  STAFFORD, DOUGLAS C.
;     TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;     TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;     NUMBER OF SEQUENCES:  32
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MEDLEN & CARROLL, LLP
;       STREET:  220 MONTGOMERY STREET, SUITE 2200
;       CITY:  SAN FRANCISCO
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,604A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/422,711
;       FILING DATE:  14-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/405,496
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  INGOLIA, DIANE E.
;       REGISTRATION NUMBER:  40,027
;       REFERENCE/DOCKET NUMBER:  OPHD-01763
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  28:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1296 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             0.8%;  Score 10;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 10
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;     APPLICANT:  WILLIAMS, JAMES A.
;     TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;     TITLE OF INVENTION:  NEUROTOXIN
;     NUMBER OF SEQUENCES:  30
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MEDLEN & CARROLL, LLP
;       STREET:  220 MONTGOMERY STREET, SUITE 2200
;       CITY:  SAN FRANCISCO
;       STATE:  CALIFORNIA
;       COUNTRY:  USA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/405,496A
;       FILING DATE:  16-MAR-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
```

```
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             0.8%;  Score 10;  DB 2;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 11
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             0.8%;  Score 10;  DB 3;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 12
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             0.8%;  Score 10;  DB 4;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        959 NNEYTIINCM 968


RESULT 13
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-22

  Query Match             0.7%;  Score 9;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        621 VPYIGLALN 629


RESULT 14
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             0.7%;  Score 9;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        621 VPYIGLALN 629


RESULT 15
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
```

; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 0.7%; Score 9; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 VPYIGLALN 611
|||||||||
Db 621 VPYIGLALN 629

Search completed: September 10, 2004, 05:17:07
Job time : 21.5272 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 63.332 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 1252
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1252	100.0	1252	12	US-10-272-898-52	Sequence 52, Appl
2	1252	100.0	1252	15	US-10-354-774-52	Sequence 52, Appl
3	1252	100.0	1252	15	US-10-271-012-52	Sequence 52, Appl
4	1252	100.0	1252	15	US-10-452-024-128	Sequence 128, App
5	1252	100.0	1252	16	US-10-729-122-52	Sequence 52, Appl
6	1252	100.0	1252	16	US-10-729-039-52	Sequence 52, Appl
7	804	64.2	804	13	US-10-011-588-37	Sequence 37, Appl
8	636	50.8	1251	15	US-10-452-024-127	Sequence 127, App
9	562	44.9	1251	15	US-10-452-024-129	Sequence 129, App
10	519	41.5	1250	12	US-10-272-898-50	Sequence 50, Appl
11	519	41.5	1250	15	US-10-354-774-50	Sequence 50, Appl
12	519	41.5	1250	15	US-10-271-012-50	Sequence 50, Appl
13	519	41.5	1250	16	US-10-729-122-50	Sequence 50, Appl
14	519	41.5	1250	16	US-10-729-039-50	Sequence 50, Appl
15	448	35.8	449	10	US-09-910-186A-14	Sequence 14, Appl
16	426	34.0	452	12	US-10-272-898-56	Sequence 56, Appl
17	426	34.0	452	15	US-10-354-774-56	Sequence 56, Appl
18	426	34.0	452	15	US-10-271-012-56	Sequence 56, Appl
19	426	34.0	452	16	US-10-729-122-56	Sequence 56, Appl
20	426	34.0	452	16	US-10-729-039-56	Sequence 56, Appl
21	423	33.8	427	13	US-10-011-588-35	Sequence 35, Appl
22	422	33.7	422	13	US-10-011-588-13	Sequence 13, Appl
23	418	33.4	1251	15	US-10-452-024-5	Sequence 5, Appli
24	418	33.4	1251	15	US-10-452-024-124	Sequence 124, App
25	418	33.4	1252	15	US-10-452-024-125	Sequence 125, App
26	418	33.4	1252	15	US-10-205-516-10	Sequence 10, Appl
27	418	33.4	1255	15	US-10-452-024-126	Sequence 126, App
28	381	30.4	382	10	US-09-910-186A-28	Sequence 28, Appl
29	236	18.8	438	15	US-10-452-024-147	Sequence 147, App
30	229	18.3	252	15	US-10-452-024-155	Sequence 155, App
31	227	18.1	1262	15	US-10-205-516-24	Sequence 24, Appl
32	140	11.2	451	12	US-10-272-898-54	Sequence 54, Appl
33	140	11.2	451	15	US-10-354-774-54	Sequence 54, Appl
34	140	11.2	451	15	US-10-271-012-54	Sequence 54, Appl
35	140	11.2	451	16	US-10-729-122-54	Sequence 54, Appl
36	140	11.2	451	16	US-10-729-039-54	Sequence 54, Appl
37	107	8.5	438	15	US-10-452-024-143	Sequence 143, App
38	52	4.2	252	15	US-10-452-024-148	Sequence 148, App
39	20	1.6	366	15	US-10-452-024-173	Sequence 173, App
40	20	1.6	369	15	US-10-452-024-174	Sequence 174, App
41	20	1.6	408	10	US-09-910-186A-30	Sequence 30, Appl
42	20	1.6	858	13	US-10-011-588-41	Sequence 41, Appl
43	20	1.6	1268	15	US-10-452-024-156	Sequence 156, App
44	20	1.6	1274	12	US-10-272-898-71	Sequence 71, Appl
45	20	1.6	1274	15	US-10-354-774-71	Sequence 71, Appl

ALIGNMENTS

```
RESULT 1
US-10-272-898-52
; Sequence 52, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
```

```
;            LENGTH: 1252 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-272-898-52

  Query Match            100.0%;  Score 1252;  DB 12;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 2
US-10-354-774-52
; Sequence 52, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1252 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-354-774-52

  Query Match            100.0%;  Score 1252;  DB 15;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
```

```
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
```

RESULT 3
US-10-271-012-52
; Sequence 52, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-271-012-52

Query Match 100.0%; Score 1252; DB 15; Length 1252;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
```

```
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252

RESULT 4
US-10-452-024-128
; Sequence 128, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 128
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-128

  Query Match           100.0%;  Score 1252;  DB 15;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
```

```
Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
```

RESULT 5
US-10-729-122-52
; Sequence 52, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-729-122-52

Query Match 100.0%; Score 1252; DB 16; Length 1252;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
```

```
Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 6
US-10-729-039-52
; Sequence 52, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1252 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-729-039-52

  Query Match           100.0%;  Score 1252;  DB 16;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
```

```
Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 7
US-10-011-588-37
; Sequence 37, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
;  APPLICANT: Smith, Leonard
;  APPLICANT:  Jensen, Melody
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 804
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:36
US-10-011-588-37

  Query Match             64.2%;  Score 804;  DB 13;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  804;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSIL 804
              ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSIL 804


RESULT 8
US-10-452-024-127
; Sequence 127, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 127
;   LENGTH: 1251
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-127

  Query Match             50.8%;  Score 636;  DB 15;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  636;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
```

```
Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEF 636
              |||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEF 636
```

RESULT 9
US-10-452-024-129
; Sequence 129, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 129
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-129

Query Match 44.9%; Score 562; DB 15; Length 1251;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 762; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        199 MNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGT 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 MNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGT 258

Qy        259 DLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYS 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 DLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYS 318

Qy        319 VNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN 378
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        319 VNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN 378

Qy        379 NLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVA 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 NLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVA 438

Qy        439 SENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYI 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 SENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYI 498

Qy        499 PKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSE 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 PKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSE 558

Qy        559 FINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQK 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 FINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQK 618

Qy        619 GNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW 678

Qy        679 KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKY 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKY 738

Qy        739 DIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYII 798
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        739 DIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYII 798

Qy        799 QHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYK 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 QHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYK 858

Qy        859 NDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSIS 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        859 NDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSIS 918

Qy        919 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWT 962
              |||||||||||||||||||||||||||||||||||||||||||||
Db        919 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWT 962
```

RESULT 10
US-10-272-898-50
; Sequence 50, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 50:

```
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1250 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-272-898-50

  Query Match             41.5%;  Score 519;  DB 12;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 11
US-10-354-774-50
; Sequence 50, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1250 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-354-774-50

  Query Match             41.5%;  Score 519;  DB 15;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 12
US-10-271-012-50
; Sequence 50, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
```

```
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/271,012
;            FILING DATE: 15-Oct-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1250 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-271-012-50

  Query Match             41.5%;  Score 519;  DB 15;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
```

```
RESULT 13
US-10-729-122-50
; Sequence 50, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,122
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1250 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-729-122-50

  Query Match             41.5%;  Score 519;  DB 16;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 14
US-10-729-039-50
; Sequence 50, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-729-039-50

  Query Match           41.5%;  Score 519;  DB 16;  Length 1250;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  519;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 15
US-09-910-186A-14
; Sequence 14, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 449
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-09-910-186A-14

  Query Match           35.8%;  Score 448;  DB 10;  Length 449;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        805 GESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVD 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          2 GESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVD 61

Qy        865 TSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIP 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIP 121

Qy        925 NYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGIS 984
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 NYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGIS 181

Qy        985 DYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGI 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 DYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGI 241

Qy       1045 RYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS 1104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 RYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS 301

Qy       1105 TLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYA 1164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYA 361

Qy       1165 DTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTW 1224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 DTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTW 421

Qy       1225 YYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||
Db        422 YYTHMRDHTNSNGCFWNFISEEHGWQEK 449
```

Search completed: September 10, 2004, 05:52:42
Job time : 66.332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 17.3447 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 1252
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1252	100.0	1252	2	S21178	botulinum neurotox
2	418	33.4	1251	2	JH0256	botulinum neurotox
3	266	21.2	367	2	S48106	neurotoxin type E
4	58	4.6	122	2	A53878	type E neurotoxin
5	20	1.6	366	2	S48110	neurotoxin type F
6	20	1.6	369	2	S48109	neurotoxin type F
7	20	1.6	1268	2	S33411	botulinum neurotox
8	20	1.6	1274	2	I40813	neurotoxin type F
9	12	1.0	1297	2	S39791	neurotoxin - Clost
10	10	0.8	13	2	S08575	botulinum neurotox
11	10	0.8	1296	1	BTCLAB	bontoxilysin (EC 3
12	10	0.8	1296	2	I40645	botulinum neurotox
13	9	0.7	1291	1	A48940	bontoxilysin (EC 3
14	9	0.7	1291	2	I40631	non-proteolytic bo
15	8	0.6	162	2	G70005	hypothetical prote
16	8	0.6	372	2	T06745	hypothetical prote
17	8	0.6	437	2	C95393	probable transport
18	8	0.6	440	2	T51556	sugar transporter-
19	8	0.6	456	2	A83264	hypothetical prote
20	8	0.6	535	2	T37508	hypothetical prote
21	8	0.6	769	2	S50966	probable membrane
22	8	0.6	773	2	C84554	hypothetical prote
23	8	0.6	1064	2	S52687	serine/threonine-s
24	7	0.6	42	2	D81005	hypothetical prote
25	7	0.6	65	2	AF3177	conserved hypothet
26	7	0.6	68	2	H85826	hypothetical prote
27	7	0.6	73	2	H71149	hypothetical prote
28	7	0.6	73	2	S10650	hypothetical prote
29	7	0.6	95	2	C86685	prophage pi1 prote
30	7	0.6	95	2	B86757	prophage pi2 prote
31	7	0.6	95	2	T13259	holin protein homo
32	7	0.6	99	2	D86797	prophage pi3 prote
33	7	0.6	112	2	D95269	hypothetical prote
34	7	0.6	117	2	H90317	conserved hypothet
35	7	0.6	118	2	T06659	histone H2A.T6G15.
36	7	0.6	118	2	C97927	hypothetical prote
37	7	0.6	118	2	B95058	transcription regu
38	7	0.6	119	2	S53460	hypothetical prote
39	7	0.6	123	2	A39266	heat shock protein
40	7	0.6	123	2	B71665	hypothetical prote
41	7	0.6	123	2	C97814	hypothetical prote
42	7	0.6	123	2	E84752	hypothetical prote
43	7	0.6	124	2	AE0480	hypothetical prote
44	7	0.6	138	2	T02889	probable carbamoyl
45	7	0.6	141	2	H64909	ydfO protein - Esc

ALIGNMENTS

RESULT 1
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin,
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin

```
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match            100.0%;  Score 1252;  DB 2;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 2
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.,
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxir
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release a
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match            33.4%;  Score 418;  DB 2;  Length 1251;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches  518;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
```

```
Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
```

RESULT 3
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   21.2%; Score 266; DB 2; Length 367;
  Best Local Similarity 99.7%; Pred. No. 2.2e-269;
  Matches 366; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 795
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 NKYDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
              |||||||
Db        361 NYGNANG 367
```

RESULT 4
A53878
type E neurotoxin - Clostridium botulinum (fragments)
C;Species: Clostridium botulinum
C;Date: 27-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C;Accession: A53878
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 11, 255-264, 1992
A;Title: Pepsin fragmentation of botulinum type E neurotoxin: isolation and characterizat
A;Reference number: A53878; MUID:93000392; PMID:1388670
A;Contents: type E, E-43, Alaska
A;Accession: A53878
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-122 <GIM>
A;Note: sequence modified after extraction from NCBI backbone
A;Note: sequence extracted from NCBI backbone (NCBIP:115403, NCBIP:115404)
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   4.6%; Score 58; DB 2; Length 122;
  Best Local Similarity   100.0%; Pred. No. 3.1e-52;
  Matches   58; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy       1116 LLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK 1173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 LLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK 119
```

RESULT 5
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   1.6%; Score 20; DB 2; Length 366;
  Best Local Similarity   100.0%; Pred. No. 4.7e-12;
  Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db         78 INTQFNKRKEQMYQALQNQV 97
```

RESULT 6
S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

```
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin

  Query Match             1.6%;  Score 20;  DB 2;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db         78 INTQFNKRKEQMYQALQNQV 97


RESULT 7
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             1.6%;  Score 20;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        702 INTQFNKRKEQMYQALQNQV 721


RESULT 8
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             1.6%;  Score 20;  DB 2;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        711 INTQFNKRKEQMYQALQNQV 730


RESULT 9
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             1.0%;  Score 12;  DB 2;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0.0034;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        205 DPALTLMHELIH 216
              ||||||||||||
Db        223 DPALTLMHELIH 234


RESULT 10
S08575
botulinum neurotoxin type E - Clostridium botulinum (strain Alaska E-43) (fragment)
C;Species: Clostridium botulinum
A;Variety: strain Alaska E-43
C;Date: 19-Mar-1997 #sequence_revision 21-Nov-1998 #text_change 21-Nov-1998
C;Accession: S08575
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S08575
A;Molecule type: protein
A;Residues: 1-13 <SCH>
A;Experimental source: strain Alaska E-43
F;1-13/Product: botulinum neurotoxin E light chain (fragment) #status predicted <LIG>

  Query Match             0.8%;  Score 10;  DB 2;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 0.0068;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 PKINSFNYND 11
              ||||||||||
Db          1 PKINSFNYND 10


RESULT 11
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
```

A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

  Query Match 0.8%; Score 10; DB 1; Length 1296;
  Best Local Similarity 100.0%; Pred. No. 0.43;
  Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 934 NNEYTIINCM 943
||||||||||
Db 959 NNEYTIINCM 968

RESULT 12
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A s
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 0.8%; Score 10; DB 2; Length 1296;
  Best Local Similarity 100.0%; Pred. No. 0.43;
  Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 946 NNSGWKVSLN 955
||||||||||
Db 970 NNSGWKVSLN 979

RESULT 13
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publicat
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisior
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus tox
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match 0.7%; Score 9; DB 1; Length 1291;
  Best Local Similarity 100.0%; Pred. No. 4.7;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 VPYIGLALN 611
 |||||||||
Db 621 VPYIGLALN 629

RESULT 14
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic si
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match 0.7%; Score 9; DB 2; Length 1291;
  Best Local Similarity 100.0%; Pred. No. 4.7;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 VPYIGLALN 611
 |||||||||
Db 621 VPYIGLALN 629

RESULT 15
G70005
hypothetical protein yuaE - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: G70005
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berterc
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G70005
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-162 <KUN>
A;Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB15081.1; PID:e1185976;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yuaE

  Query Match 0.6%; Score 8; DB 2; Length 162;
  Best Local Similarity 100.0%; Pred. No. 8.1;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 482 GLSDEKLN 489
 ||||||||
Db 17 GLSDEKLN 24

Search completed: September 10, 2004, 05:14:49
Job time : 18.4447 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 10.8205 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 1252
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	562	44.9	1250	1	BXE_CLOBO	Q00496 clostridium
2	418	33.4	1250	1	BXE_CLOBU	P30995 clostridium
3	20	1.6	1274	1	BXF_CLOBO	P30996 clostridium
4	12	1.0	1296	1	BXG_CLOBO	Q60393 clostridium
5	10	0.8	1295	1	BXA1_CLOBO	P10845 clostridium
6	10	0.8	1295	1	BXA2_CLOBO	Q45894 clostridium
7	9	0.7	550	1	RPOA_CHLRE	Q8huh2 chlamydomon
8	9	0.7	1290	1	BXB_CLOBO	P10844 clostridium
9	8	0.6	838	1	HIS2_CANAL	O74712 candida alb
10	8	0.6	980	1	SYN_DROME	Q24546 drosophila
11	8	0.6	1064	1	KIN1_YEAST	P13185 saccharomyc
12	7	0.6	73	1	TRM1_PYRWO	P20300 pyrococcus
13	7	0.6	123	1	Y600_RICPR	Q9zcv8 rickettsia
14	7	0.6	123	1	Y915_RICCN	Q9akn7 rickettsia
15	7	0.6	136	1	YDFO_ECOLI	P76156 escherichia
16	7	0.6	160	1	RL11_METTH	O27715 methanobact
17	7	0.6	161	1	LI52_CAEEL	Q10120 caenorhabdi
18	7	0.6	180	1	PUR6_ARCFU	O28997 archaeoglob
19	7	0.6	189	1	E2BE_CANAL	P87163 candida alb
20	7	0.6	192	1	Y3_SOCMV	P15633 soybean chl
21	7	0.6	193	1	Y455_METJA	Q57897 methanococc
22	7	0.6	211	1	SN23_HUMAN	O00161 homo sapien
23	7	0.6	218	1	KAD_CLOPE	Q8xhu4 clostridium
24	7	0.6	218	1	RR3_LOTJA	Q9bbp8 lotus japon
25	7	0.6	230	1	AG11_ARATH	Q38836 arabidopsis
26	7	0.6	240	1	MYF6_XENLA	Q92020 xenopus lae
27	7	0.6	243	1	FRHG_METVO	Q00393 methanococc
28	7	0.6	244	1	ARC3_CBCP	Q00901 clostridium
29	7	0.6	246	1	TRMD_LACPL	Q88wj2 lactobacill
30	7	0.6	251	1	RECO_STRCO	Q912h3 streptomyce
31	7	0.6	256	1	TPIS_WIGBR	Q8d2t0 wiggleswort
32	7	0.6	274	1	TPM1_METEN	Q25456 metapenaeus
33	7	0.6	274	1	TPM1_PANST	O61379 panulirus s
34	7	0.6	304	1	LST_HAEIN	Q48211 haemophilus
35	7	0.6	315	1	RLUC_BUCBP	Q89ah2 buchnera ap
36	7	0.6	325	1	GPR_CLOPE	Q8xis3 clostridium
37	7	0.6	337	1	HNR_ECOLI	P37055 escherichia
38	7	0.6	351	1	UXUA_CLOAB	Q97jf4 clostridium
39	7	0.6	357	1	TCSA_LISIN	Q92bw7 listeria in
40	7	0.6	357	1	TCSA_LISMO	Q48754 listeria mo
41	7	0.6	359	1	ALF_HAEIN	P44429 haemophilus
42	7	0.6	381	1	ADHX_MAIZE	P93629 zea mays (m
43	7	0.6	381	1	TRM1_PYRFU	P81554 pyrococcus
44	7	0.6	383	1	TRM1_PYRAB	Q9v1p3 pyrococcus
45	7	0.6	397	1	FTSW_ENTHR	Q47866 enterococcu

ALIGNMENTS

RESULT 1
BXE_CLOBO
ID BXE_CLOBO STANDARD; PRT; 1250 AA.
AC Q00496;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beluga;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinal neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [4]
RP SEQUENCE OF 1-13.
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [5]
RP SEQUENCE OF 419-426.
RX MEDLINE=90344918; PubMed=2116911;
RA Gimenez J.A., Dasgupta B.R.;
RT "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RT neurotoxin.";
RL Biochimie 72:213-217(1990).
RN [6]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,

```
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             44.9%;  Score 562;  DB 1;  Length 1250;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches  762;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       199 MNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGT 258
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       198 MNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGT 257

Qy       259 DLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYS 318
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       258 DLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYS 317

Qy       319 VNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN 378
            ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       318 VNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN 377

Qy       379 NLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVA 438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       378 NLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVA 437

Qy       439 SENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYI 498
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       438 SENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYI 497

Qy       499 PKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSE 558
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       498 PKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSE 557

Qy       559 FINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQK 618
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       558 FINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQK 617

Qy       619 GNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW 678
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       618 GNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW 677

Qy       679 KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKY 738
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       678 KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKY 737

Qy       739 DIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYII 798
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       738 DIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYII 797

Qy       799 QHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYK 858
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       798 QHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYK 857

Qy       859 NDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSIS 918
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       858 NDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSIS 917

Qy       919 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWT 962
            |||||||||||||||||||||||||||||||||||||||||||||
Db       918 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWT 961


RESULT 2
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
```

```
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS.NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             33.4%;  Score 418;  DB 1;  Length 1250;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  518;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 254

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 314

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 374

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 434

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 494

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 554

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        555 SSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 614

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 674

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        675 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 713


RESULT 3
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
```

```
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             1.6%;  Score 20;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 3e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        711 INTQFNKRKEQMYQALQNQV 730
```

```
RESULT 4
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             1.0%;  Score 12;  DB 1;  Length 1296;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0.0033;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        205 DPALTLMHELIH 216
              ||||||||||||
Db        222 DPALTLMHELIH 233


RESULT 5
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
```

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             0.8%;  Score 10;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.34;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 NNEYTIINCM 943
              ||||||||||
Db        958 NNEYTIINCM 967
```

```
RESULT 6
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
```

```
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             0.8%;  Score 10;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.34;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        946 NNSGWKVSLN 955
              ||||||||||
Db        969 NNSGWKVSLN 978


RESULT 7
RPOA_CHLRE
ID   RPOA_CHLRE     STANDARD;      PRT;   550 AA.
AC   Q8HUH2;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (PEP) (Plastid-
DE   encoded RNA polymerase alpha subunit) (RNA polymerase alpha subunit).
GN   RPOA.
OS   Chlamydomonas reinhardtii.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A., IDENTIFICATION, COMPLETE PLASTID GENOME, AND
RP   EXPRESSION OF MRNA.
RX   MEDLINE=22305394; PubMed=12417694;
RA   Maul J.E., Lilly J.W., Cui L., dePamphilis C.W., Miller W.,
RA   Harris E.H., Stern D.B.;
RT   "The Chlamydomonas reinhardtii plastid chromosome: islands of genes in
RT   a sea of repeats.";
RL   Plant Cell 14:2659-2679(2002).
CC   -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC       of DNA into RNA using the four ribonucleoside triphosphates as
CC       substrates.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: In chloroplasts the minimal PEP RNA polymerase is
CC       composed of four subunits: alpha, beta, beta', and beta'' (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- DOMAIN: The N-terminal domain is essential for RNAP assembly and
CC       basal transcription, whereas the C-terminal domain is involved in
CC       interaction with transcriptional regulators and with upstream
CC       promoter elements (By similarity).
CC   -!- SIMILARITY: Belongs to the RNA polymerase alpha chain family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF541867; AAN17816.1; -.
DR   EMBL; BK000554; DAA00934.1; -.
DR   HAMAP; MF_00059; atypical; 1.
DR   InterPro; IPR009025; RBP11-like_RNApo.
DR   InterPro; IPR001700; RNA_polA_bac_org.
DR   Pfam; PF01000; RNA_pol_A_bac; 1.
DR   ProDom; PD001179; RNA_polA_bac_org; 1.
DR   SMART; SM00662; RPOLD; 1.
KW   Transcription; Transferase; DNA-directed RNA polymerase; Chloroplast.
FT   DOMAIN        1    333       ALPHA N-TERMINAL DOMAIN (ALPHA-NTD) (BY
FT                                SIMILARITY).
FT   DOMAIN      185    258       INSERT.
FT   DOMAIN      378    550       ALPHA C-TERMINAL DOMAIN (ALPHA-CTD) (BY
FT                                SIMILARITY).
SQ   SEQUENCE   550 AA;  62290 MW;  931F910FF6359948 CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        650 IKSFLGSSD 658
              |||||||||
Db        447 IKSFLGSSD 455


RESULT 8
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
```

```
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match              0.7%;  Score 9;  DB 1;  Length 1290;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 VPYIGLALN 611
              |||||||||
Db        620 VPYIGLALN 628


RESULT 9
HIS2_CANAL
ID   HIS2_CANAL     STANDARD;      PRT;   838 AA.
AC   O74712;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Histidine biosynthesis trifunctional protein [Includes:
DE   Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19); Phosphoribosyl-ATP
DE   pyrophosphohydrolase (EC 3.6.1.31); Histidinol dehydrogenase
DE   (EC 1.1.1.23) (HDH)].
GN   HIS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 64385 / 1001;
RX   MEDLINE=98451822; PubMed=9778800;
RA   Navarro-Garcia F., Perez-Diaz R., Negredo A., Pla J., Nombela C.;
RT   "Cloning and sequence of a 3.835 kbp DNA fragment containing the HIS4
RT   gene and a fragment of a PEX5-like gene from Candida albicans.";
RL   Yeast 14:1147-1157(1998).
CC   -!- CATALYTIC ACTIVITY: 5-phosphoribosyl-AMP + H(2)O = 5-(5-phospho-D-
CC       ribosylaminoformimino)-1-(5-phospho-ribosyl) imidazole-4-
CC       carboxamide.
CC   -!- CATALYTIC ACTIVITY: 5-phosphoribosyl-ATP + H(2)O = 5-
CC       phosphoribosyl-AMP + diphosphate.
CC   -!- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2)O = L-histidine
CC       + 2 NADH.
CC   -!- COFACTOR: Binds 1 zinc ion (By similarity).
CC   -!- PATHWAY: Histidine biosynthesis; second step.
CC   -!- PATHWAY: Histidine biosynthesis; third step.
CC   -!- PATHWAY: Histidine biosynthesis; ninth (last) step.
CC   -!- SIMILARITY: In the C-terminal section; belongs to the histidinol
CC       dehydrogenase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ003115; CAA05871.1; -.
DR   InterPro; IPR001692; Histidinol_dh.
DR   InterPro; IPR002496; PRA-CH.
DR   InterPro; IPR008179; PRA-PH.
DR   InterPro; IPR008178; Pra_PH/CH.
DR   Pfam; PF00815; Histidinol_dh; 1.
DR   Pfam; PF01502; PRA-CH; 1.
```

```
DR   Pfam; PF01503; PRA-PH; 1.
DR   PRINTS; PR00083; HOLDHDRGNASE.
DR   ProDom; PD002680; Histidinol_dh; 1.
DR   ProDom; PD002610; PRA-CH; 1.
DR   ProDom; PD002611; Pra_PH/CH; 1.
DR   TIGRFAMs; TIGR00069; hisD; 1.
DR   PROSITE; PS00611; HISOL_DEHYDROGENASE; 1.
KW   Histidine biosynthesis; Multifunctional enzyme; Hydrolase;
KW   Oxidoreductase; NAD; Metal-binding; Zinc.
FT   DOMAIN        1    271       PHOSPHORIBOSYL-AMP CYCLOHYDROLASE.
FT   DOMAIN      272    360       PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE.
FT   DOMAIN      361    838       HISTIDINOL DEHYDROGENASE.
FT   ACT_SITE    729    729       BY SIMILARITY.
FT   ACT_SITE    730    730       BY SIMILARITY.
FT   METAL       660    660       ZINC (BY SIMILARITY).
FT   METAL       663    663       ZINC (BY SIMILARITY).
FT   METAL       764    764       ZINC (BY SIMILARITY).
FT   METAL       823    823       ZINC (BY SIMILARITY).
SQ   SEQUENCE   838 AA;  91835 MW;  BFA0E97E9CA62C03 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 838;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        482 GLSDEKLN 489
              ||||||||
Db        675 GLSDEKLN 682


RESULT 10
SYN_DROME
ID   SYN_DROME      STANDARD;      PRT;   980 AA.
AC   Q24546; Q24545;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Synapsin.
GN   SYN OR SYN1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS SYN-S AND SYN1-RT).
RC   STRAIN=Berlin; TISSUE=Head;
RX   MEDLINE=96209149; PubMed=8627354;
RA   Klagges B.R.E., Heimbeck G., Godenschwege T.A., Hofbauer A.,
RA   Pflugfelder G.O., Reifegerste R., Reisch D., Schaupp M., Buchner S.,
RA   Buchner E.;
RT   "Invertebrate synapsins: a single gene codes for several isoforms in
RT   Drosophila.";
RL   J. Neurosci. 16:3154-3165(1996).
CC   -!- SUBCELLULAR LOCATION: Synapse.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Syn1-RT;
CC         IsoId=Q24546-1; Sequence=Displayed;
CC       Name=Syn-S;
CC         IsoId=Q24546-2; Sequence=VSP_006324;
CC   -!- TISSUE SPECIFICITY: Widely expressed in the nervous system.
CC   -!- SIMILARITY: Belongs to the synapsin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95453; CAA64723.1; -.
DR   EMBL; X95453; CAA64722.1; -.
DR   HSSP; P17599; 1AUX.
DR   FlyBase; FBgn0004575; Syn.
DR   InterPro; IPR001359; Synapsin.
DR   Pfam; PF02078; Synapsin; 1.
DR   Pfam; PF02750; Synapsin_C; 1.
DR   PRINTS; PR01368; SYNAPSIN.
DR   PROSITE; PS00415; SYNAPSIN_1; FALSE_NEG.
DR   PROSITE; PS00416; SYNAPSIN_2; FALSE_NEG.
KW   Synapse; Neurone; Alternative splicing.
FT   DOMAIN        2      7       POLY-PRO.
FT   DOMAIN      668    680       POLY-SER.
FT   DOMAIN      692    695       POLY-PRO.
FT   DOMAIN      698    703       POLY-PRO.
FT   VARSPLIC    539    980       Missing (in isoform Syn-S).
FT                                /FTId=VSP_006324.
SQ   SEQUENCE   980 AA;  102799 MW;  E6037A2A4804F944 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 980;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        864 DTSGYDSN 871
              ||||||||
Db        823 DTSGYDSN 830


RESULT 11
KIN1_YEAST
ID   KIN1_YEAST     STANDARD;      PRT;  1064 AA.
AC   P13185; Q04606;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Protein kinase KIN1 (EC 2.7.1.-).
GN   KIN1 OR YDR122W OR YD9727.17.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87317589; PubMed=2957690;
RA   Levin D.E., Hammond C.I., Ralston R.O., Bishop J.M.;
RT   "Two yeast genes that encode unusual protein kinases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:6035-6039(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Murphy L., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
RA   Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This protein is probably a serine/threonine protein
CC       kinase.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. NIM1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69017; AAA34722.1; -.
DR   EMBL; Z48758; CAA88675.1; -.
DR   PIR; S52687; S52687.
DR   HSSP; Q63450; 1A06.
DR   GermOnline; 140613; -.
DR   SGD; S0002529; KIN1.
DR   GO; GO:0005624; C:membrane fraction; IDA.
DR   GO; GO:0004672; F:protein kinase activity; IDA.
DR   InterPro; IPR001772; Kinase_Cterm.
```

```
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF02149; KA1; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN      120    398       PROTEIN KINASE.
FT   NP_BIND     126    134       ATP (BY SIMILARITY).
FT   BINDING     149    149       ATP (BY SIMILARITY).
FT   ACT_SITE    269    269       BY SIMILARITY.
FT   CONFLICT     25     25       S -> R (IN REF. 1).
FT   CONFLICT    453    453       T -> H (IN REF. 1).
FT   CONFLICT    455    455       V -> G (IN REF. 1).
FT   CONFLICT    718    718       A -> R (IN REF. 1).
FT   CONFLICT    920    921       NI -> IN (IN REF. 1).
FT   CONFLICT    976    976       T -> A (IN REF. 1).
FT   CONFLICT    979    980       SI -> T (IN REF. 1).
FT   CONFLICT    984    985       KT -> NS (IN REF. 1).
SQ   SEQUENCE   1064 AA;  120070 MW;  E4104B84A8E45C36 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 1064;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        796 YIIQHGSI 803
              ||||||||
Db        234 YIIQHGSI 241


RESULT 12
TRM1_PYRWO
ID   TRM1_PYRWO     STANDARD;      PRT;    73 AA.
AC   P20300;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   N(2),N(2)-dimethylguanosine tRNA methyltransferase (EC 2.1.1.32)
DE   (tRNA(guanine-26,N(2)-N(2)) methyltransferase) (tRNA 2,2-
DE   dimethylguanosine-26 methyltransferase)
DE   (tRNA(m(2,2)G26)dimethyltransferase) (Fragment).
GN   TRM1.
OS   Pyrococcus woesei.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 3773;
RX   MEDLINE=90330536; PubMed=2165475;
RA   Zwickl P., Fabry S., Bogedain C., Haas A., Hensel R.;
RT   "Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic
RT   archaebacterium Pyrococcus woesei: characterization of the enzyme,
RT   cloning and sequencing of the gene, and expression in Escherichia
RT   coli.";
RL   J. Bacteriol. 172:4329-4338(1990).
RN   [2]
RP   SIMILARITY.
RX   MEDLINE=95206934; PubMed=7899076;
RA   Ouzounis C., Kyrpides N., Sander C.;
RT   "Novel protein families in archaean genomes.";
RL   Nucleic Acids Res. 23:565-570(1995).
CC   -!- FUNCTION: Dimethylates a single guanine residue at position 26 of
CC       a number of tRNAs using S-adenosyl-L-methionine as donor of the
CC       methyl groups (By similarity).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC       homocysteine + tRNA containing N(2)-methylguanine.
CC   -!- SIMILARITY: Belongs to the TRM1 family.
DR   PIR; S10650; S10650.
DR   HAMAP; MF_00290; -; 1.
DR   InterPro; IPR002905; TRM.
DR   Pfam; PF02005; TRM; 1.
KW   Transferase; Methyltransferase; tRNA processing.
FT   NON_TER      73     73
SQ   SEQUENCE   73 AA;  8094 MW;  A5413AEFDF00C00C CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1088 LLNVLKP 1094
              |||||||
Db         44 LLNVLKP 50


RESULT 13
Y600_RICPR
ID   Y600_RICPR     STANDARD;      PRT;   123 AA.
AC   Q9ZCV8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein RP600.
GN   RP600.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- SIMILARITY: Contains 1 rhodanese domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235272; CAA15044.1; -.
DR   PIR; B71665; B71665.
DR   InterPro; IPR001763; Rhodanese-like.
DR   SMART; SM00450; RHOD; 1.
DR   PROSITE; PS50206; RHODANESE_3; 1.
KW   Hypothetical protein; Complete proteome.
FT   DOMAIN       17    117       RHODANESE.
SQ   SEQUENCE   123 AA;  14320 MW;  A08389F4B8ED185D CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        658 DNKNKVI 664
              |||||||
Db         41 DNKNKVI 47


RESULT 14
Y915_RICCN
ID   Y915_RICCN     STANDARD;      PRT;   123 AA.
AC   Q9AKN7;
```

```
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein RC0915.
GN   RC0915.
OS   Rickettsia conorii, and
OS   Rickettsia montana.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781, 33991;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=R.conorii; STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=R.montana;
RX   MEDLINE=21219194; PubMed=11319266;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes.";
RL   Mol. Biol. Evol. 18:829-839(2001).
CC   -!- SIMILARITY: Contains 1 rhodanese domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008645; AAL03453.1; -.
DR   EMBL; AJ293315; CAC33611.1; -.
DR   PIR; C97814; C97814.
DR   InterPro; IPR001763; Rhodanese-like.
DR   SMART; SM00450; RHOD; 1.
DR   PROSITE; PS50206; RHODANESE_3; 1.
KW   Hypothetical protein; Complete proteome.
FT   DOMAIN       17    117       RHODANESE.
SQ   SEQUENCE   123 AA;  14408 MW;  7108705FCBCF1480 CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        658 DNKNKVI 664
              |||||||
Db         41 DNKNKVI 47


RESULT 15
YDFO_ECOLI
ID   YDFO_ECOLI     STANDARD;      PRT;   136 AA.
AC   P76156;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ydfO.
GN   YDFO OR B1549.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: STRONG, TO E.COLI YBCV AND YCGX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000252; AAC74622.1; ALT_INIT.
DR   EcoGene; EG13827; ydfO.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   136 AA;  16497 MW;  554B812E2DD7B33C CRC64;

  Query Match             0.6%;  Score 7;  DB 1;  Length 136;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        413 KNIVSVK 419
              |||||||
Db         60 KNIVSVK 66


Search completed: September 10, 2004, 05:07:18
Job time : 12.8205 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 50.9202 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 1252
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	636	50.8	1251	2	Q9K395	Q9k395 clostridium
2	418	33.4	1252	2	Q8KZM3	Q8kzm3 clostridium
3	418	33.4	1255	2	Q9FAR6	Q9far6 clostridium
4	367	29.3	367	2	Q45862	Q45862 clostridium
5	266	21.2	367	2	Q45861	Q45861 clostridium
6	173	13.8	173	2	Q9KW90	Q9kw90 clostridium
7	114	9.1	173	2	Q9KW89	Q9kw89 clostridium
8	79	6.3	173	2	Q9KW88	Q9kw88 clostridium
9	76	6.1	173	2	Q9K345	Q9k345 clostridium
10	57	4.6	60	2	Q9R5H0	Q9r5h0 clostridium
11	42	3.4	52	2	Q9R5H1	Q9r5h1 clostridium
12	41	3.3	71	2	Q7WUH7	Q7wuh7 clostridium
13	41	3.3	71	2	Q7WRP2	Q7wrp2 clostridium
14	20	1.6	1268	2	Q45851	Q45851 clostridium
15	20	1.6	1278	2	Q57236	Q57236 clostridium
16	20	1.6	1280	2	Q9ZAJ5	Q9zaj5 clostridium
17	9	0.7	73	2	Q7WUH8	Q7wuh8 clostridium
18	9	0.7	361	2	Q45848	Q45848 clostridium
19	9	0.7	361	2	Q45846	Q45846 clostridium
20	9	0.7	1165	5	Q8I341	Q8i341 plasmodium
21	9	0.7	1291	2	Q9ZAJ8	Q9zaj8 clostridium
22	9	0.7	1291	2	Q93G71	Q93g71 clostridium
23	9	0.7	1291	2	Q933K0	Q933k0 clostridium
24	9	0.7	1291	2	Q08077	Q08077 clostridium
25	9	0.7	1291	2	Q8GR96	Q8gr96 clostridium
26	8	0.6	49	2	O87711	O87711 clostridium
27	8	0.6	82	16	Q82ZU5	Q82zu5 enterococcu
28	8	0.6	120	2	Q8KJT2	Q8kjt2 proteus vul
29	8	0.6	162	16	O32078	O32078 bacillus su
30	8	0.6	197	16	Q7U7I6	Q7u7i6 synechococc
31	8	0.6	209	11	Q9D8N8	Q9d8n8 mus musculu
32	8	0.6	306	16	Q8D8A0	Q8d8a0 vibrio vuln
33	8	0.6	306	16	Q87MY6	Q87my6 vibrio para
34	8	0.6	334	10	Q8LGB6	Q8lgb6 arabidopsis
35	8	0.6	342	10	Q9M9M8	Q9m9m8 arabidopsis
36	8	0.6	347	2	Q840G4	Q840g4 clostridium
37	8	0.6	372	10	Q9SVY9	Q9svy9 arabidopsis
38	8	0.6	377	16	Q7VYG2	Q7vyg2 bordetella
39	8	0.6	391	16	Q8RCZ3	Q8rcz3 thermoanaer
40	8	0.6	394	17	Q8TUI9	Q8tui9 methanosarc
41	8	0.6	397	2	Q9S523	Q9s523 escherichia
42	8	0.6	402	16	Q81PW2	Q81pw2 bacillus an
43	8	0.6	402	16	Q81CQ8	Q81cq8 bacillus ce
44	8	0.6	437	16	Q92Y37	Q92y37 rhizobium m
45	8	0.6	440	10	Q9LFK1	Q9lfk1 arabidopsis

ALIGNMENTS

RESULT 1
Q9K395
ID Q9K395 PRELIMINARY; PRT; 1251 AA.
AC Q9K395;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 095;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 155;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA Kato H., Nakamura S., Karasawa T.;
RT "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1899;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1897;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;

```
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match           50.8%;  Score 636;  DB 2;  Length 1251;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  636;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEF 636
              ||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEF 636


RESULT 2
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
```

```
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match          33.4%;  Score 418;  DB 2;  Length 1252;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  518;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        556 SSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714


RESULT 3
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match          33.4%;  Score 418;  DB 2;  Length 1255;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  518;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        196 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 DNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTF 258

Qy        256 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 GGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASG 318

Qy        316 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 IYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGY 378

Qy        376 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 NINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELF 438

Qy        436 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 FVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQND 498

Qy        496 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFF 558

Qy        556 SSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 615
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        559 SSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNE 618

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 678

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 714
              |||||||||||||||||||||||||||||||||||||||
Db        679 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNA 717


RESULT 4
Q45862
ID   Q45862      PRELIMINARY;      PRT;   367 AA.
AC   Q45862;
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type E;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70815; CAA50146.1; -.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42854 MW;  0810595B3A865570 CRC64;

  Query Match             29.3%;  Score 367;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  367;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 735
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 795
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 975
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
              |||||||
Db        361 NYGNANG 367


RESULT 5
Q45861
ID   Q45861      PRELIMINARY;      PRT;   367 AA.
AC   Q45861;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type E;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70818; CAA50149.1; -.
DR   PIR; S48106; S48106.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42902 MW;  346A610C2FF70262 CRC64;

  Query Match             21.2%;  Score 266;  DB 2;  Length 367;
  Best Local Similarity   99.7%;  Pred. No. 1.3e-269;
  Matches  366;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 735
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 795
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NKYDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 975
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
              |||||||
Db        361 NYGNANG 367


RESULT 6
Q9KW90
ID   Q9KW90      PRELIMINARY;      PRT;   173 AA.
AC   Q9KW90;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Biwako;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=35396;
RA   Kimura B.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35396;
RA   Kmura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Real-time, quantitative PCR assay of Clostriium botulinum type E in
RT   modified-atmosphere-packaged fish.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB040125; BAB07887.2; -.
DR   EMBL; AB040127; BAB07889.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19151 MW;  41781916CF06AB1F CRC64;

  Query Match             13.8%;  Score 173;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-172;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 60

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 120

Qy        580 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGI 632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGI 173


RESULT 7
Q9KW89
ID   Q9KW89      PRELIMINARY;      PRT;   173 AA.
AC   Q9KW89;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=5545;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
DR   EMBL; AB040126; BAB07888.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19137 MW;  4178024F5BC4BF1F CRC64;

  Query Match             9.1%;  Score 114;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-110;
  Matches  114;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 60

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFV 573
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFV 114


RESULT 8
Q9KW88
ID   Q9KW88      PRELIMINARY;      PRT;   173 AA.
AC   Q9KW88;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=164-1;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
DR   EMBL; AB040128; BAB07890.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19121 MW;  42F5822DE3B01F1F CRC64;

  Query Match             6.3%;  Score 79;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 1e-73;
  Matches   79;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        495 DAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTF 554
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         36 DAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTF 95

Qy        555 FSSEFINNVNKPVQAALFV 573
              |||||||||||||||||||
Db         96 FSSEFINNVNKPVQAALFV 114


RESULT 9
Q9K345
ID   Q9K345      PRELIMINARY;      PRT;   173 AA.
AC   Q9K345; Q9KW91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Iwanai, and Tenno2;
RX   MEDLINE=20575211; PubMed=11133447;
RA   Kimura B., Kawasaki S., Nakano H., Fujii T.;
RT   "Rapid, Quantitative PCR Monitoring of Growth of Clostridium botulinum
RT   Type E in Modified-Atmosphere-Packaged Fish.";
RL   Appl. Environ. Microbiol. 67:206-216(2001).
DR   EMBL; AB040123; BAB07885.2; -.
DR   EMBL; AB040124; BAB07886.2; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
```

```
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     173    173
SQ   SEQUENCE   173 AA;  19061 MW;  5A89702DE3ACED16 CRC64;

  Query Match             6.1%;  Score 76;  DB 2;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-70;
  Matches   76;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        498 IPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSS 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 IPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSS 98

Qy        558 EFINNVNKPVQAALFV 573
              ||||||||||||||||
Db         99 EFINNVNKPVQAALFV 114


RESULT 10
Q9R5H0
ID   Q9R5H0      PRELIMINARY;      PRT;    60 AA.
AC   Q9R5H0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E neurotoxin, type E NT=16 kDa fragment (Fragment).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93000392; PubMed=1388670;
RA   Gimenez J.A., DasGupta B.R.;
RT   "Pepsin fragmentation of botulinum type E neurotoxin: isolation and
RT   characterization of 112, 48, 46, and 16 kD fragments.";
RL   J. Protein Chem. 11:255-264(1992).
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   InterPro; IPR002160; Kunitz_legume.
FT   NON_TER       1      1
FT   NON_TER      60     60
SQ   SEQUENCE   60 AA;  6865 MW;  9EA8C4CABFEAE766 CRC64;

  Query Match             4.6%;  Score 57;  DB 2;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-51;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1117 LANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK 1173
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK 57


RESULT 11
Q9R5H1
ID   Q9R5H1      PRELIMINARY;      PRT;    52 AA.
AC   Q9R5H1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E neurotoxin, type E NT=46 kDa fragment (Fragment).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93000392; PubMed=1388670;
RA   Gimenez J.A., DasGupta B.R.;
RT   "Pepsin fragmentation of botulinum type E neurotoxin: isolation and
RT   characterization of 112, 48, 46, and 16 kD fragments.";
RL   J. Protein Chem. 11:255-264(1992).
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
FT   NON_TER       1      1
FT   NON_TER      52     52
SQ   SEQUENCE   52 AA;  6122 MW;  38AE1305807188D1 CRC64;

  Query Match             3.4%;  Score 42;  DB 2;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-35;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        844 FKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNK 885
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 FKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNK 42


RESULT 12
Q7WUH7
ID   Q7WUH7      PRELIMINARY;      PRT;    71 AA.
AC   Q7WUH7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=21;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327859; AAQ16540.1; -.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER      71     71
SQ   SEQUENCE   71 AA;  8391 MW;  A1B3F4F757EF6122 CRC64;

  Query Match             3.3%;  Score 41;  DB 2;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-34;
  Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1103 DSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVR 1143
              |||||||||||||||||||||||||||||||||||||||||||
Db         25 DSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVR 65


RESULT 13
Q7WRP2
ID   Q7WRP2      PRELIMINARY;      PRT;    71 AA.
AC   Q7WRP2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type E (Fragment).
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=109, and LCL155;
RA   Franciosa G., Pourshaban M., De Luca A., Buccino A., Dallapiccola B.,
RA   Aureli P.;
RT   "Identification of botulinum neurotoxin gene types and of
RT   neurotoxigenic clostridia by denaturing HPLC (DHPLC).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY327860; AAQ16541.1; -.
DR   EMBL; AY327861; AAQ16542.1; -.
KW   Neurotoxin.
```

```
FT   NON_TER       1      1
FT   NON_TER      71     71
SQ   SEQUENCE   71 AA;  8360 MW;  BDC3E11757EF6122 CRC64;

  Query Match             3.3%;  Score 41;  DB 2;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-34;
  Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1103 DSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVR 1143
              |||||||||||||||||||||||||||||||||||||||||||
Db         25 DSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVR 65


RESULT 14
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match             1.6%;  Score 20;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        702 INTQFNKRKEQMYQALQNQV 721


RESULT 15
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match             1.6%;  Score 20;  DB 2;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        693 INTQFNKRKEQMYQALQNQV 712
              ||||||||||||||||||||
Db        712 INTQFNKRKEQMYQALQNQV 731


Search completed: September 10, 2004, 05:12:51
Job time : 51.9202 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 82.5862 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 6541
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4153	63.5	804	5	ABG69083	Abg69083 Botulinum
2	3040.5	46.5	1059	3	AAY93309	Aay93309 A mangane
3	3040.5	46.5	1084	3	AAY93312	Aay93312 A mangane
4	2728.5	41.7	858	5	ABG69085	Abg69085 Botulinum
5	2384	36.4	449	3	AAY77137	Aay77137 Synthetic
6	2384	36.4	449	4	AAB04094	Aab04094 Botulism
7	2336.5	35.7	1296	2	AAR95010	Aar95010 C. botuli
8	2324	35.5	1295	5	AAU99339	Aau99339 Clostridi
9	2267	34.7	452	2	AAW68396	Aaw68396 Clostridi
10	2230	34.1	419	4	AAB04095	Aab04095 Botulism
11	2212	33.8	422	5	ABG69071	Abg69071 Botulinum
12	2210.5	33.8	451	2	AAW68395	Aaw68395 Clostridi
13	2210	33.8	427	5	ABG69082	Abg69082 Botulinum
14	2194	33.5	1291	2	AAW68392	Aaw68392 Clostridi
15	2151.5	32.9	1315	4	AAB61169	Aab61169 Clostridi
16	2141.5	32.7	1169	2	AAW56017	Aaw56017 Recombina
17	1936	29.6	382	4	AAB04100	Aab04100 Botulism
18	1812	27.7	1067	3	AAY93307	Aay93307 A mangane
19	1812	27.7	1092	3	AAY93310	Aay93310 A mangane
20	1799	27.5	1291	2	AAY05814	Aay05814 Non-toxic
21	1785	27.3	837	3	AAY77140	Aay77140 Native bo
22	1772.5	27.1	847	4	AAB04081	Aab04081 Botulism
23	1548	23.7	1092	4	AAE07900	Aae07900 C. botuli
24	1536	23.5	1070	3	AAY93308	Aay93308 A mangane
25	1536	23.5	1095	3	AAY93311	Aay93311 A mangane
26	1526.5	23.3	858	2	AAW56018	Aaw56018 Recombina
27	1520	23.2	1032	4	AAE07901	Aae07901 C. botuli
28	1512	23.1	852	5	ABG69077	Abg69077 Botulinum
29	1510	23.1	848	4	AAB04082	Aab04082 Botulism
30	1499.5	22.9	848	5	ABG69087	Abg69087 Botulinum
31	1487.5	22.7	1192	6	AAE35711	Aae35711 BoNT/F-Hc
32	1487.5	22.7	1192	6	AAE35710	Aae35710 BoNT/F-Hc
33	1476.5	22.6	408	4	AAB04101	Aab04101 Botulism
34	1468.5	22.5	979	6	AAE35713	Aae35713 BoNT/F-Hc
35	1466	22.4	887	4	AAE07892	Aae07892 Modified
36	1465	22.4	862	4	AAE07890	Aae07890 Modified
37	1459	22.3	685	4	AAE07893	Aae07893 Modified
38	1457.5	22.3	431	2	AAW09014	Aaw09014 Immunogen
39	1457.5	22.3	432	4	AAB04103	Aab04103 Botulism
40	1457.5	22.3	432	4	AAB04096	Aab04096 Botulism
41	1457.5	22.3	645	4	AAE07894	Aae07894 Modified
42	1457.5	22.3	645	6	AAE35692	Aae35692 DipT HN d
43	1457.5	22.3	657	6	AAE35693	Aae35693 BoNT/F-Hc
44	1457.5	22.3	657	6	AAE35694	Aae35694 BoNT/F-Hc
45	1440.5	22.0	432	3	AAY77138	Aay77138 Synthetic

ALIGNMENTS

RESULT 1
ABG69083
ID ABG69083 standard; protein; 804 AA.
XX
AC ABG69083;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #17.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98553.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 152-154; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

```
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 804 AA;

  Query Match             63.5%;  Score 4153;  DB 5;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-221;
  Matches  804;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSIL 804
              ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSIL 804


RESULT 2
AAY93309
ID   AAY93309 standard; protein; 1059 AA.
XX
AC   AAY93309;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 48-51; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC   be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype F. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a domain
CC   that translocates the SOD of the composition into the neuronal cell.
CC   After translocation, the linker is cleaved to release the SOD. The
CC   composition is useful for treating neuronal diseases caused or augmented
CC   by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC   disease, Huntington's disease and motor neurone diseases
XX
SQ   Sequence 1059 AA;

  Query Match             46.5%;  Score 3040.5;  DB 3;  Length 1059;
  Best Local Similarity   68.2%;  Pred. No. 2.2e-159;
  Matches  572;  Conservative  133;  Mismatches  117;  Indels   17;  Gaps    7;

Qy        425 ICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLS 484
              :|| :|| |||||||||:|||:::|||||||||||   |||| |:||:|||::|||: | :|
Db        225 LCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLNNNYRNNLDEVILDYNSETIPQIS 284

Qy        485 DEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTA 544
              :: ||  :|:|:|:|:|||||||:||:|:| :|||||||  ||||||||  |::|||||||||
Db        285 NQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFFYLHAQKVPEGETNISLTSSIDTA 344

Qy        545 LLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVP 604
              | |: ::|||||||||| :|||| ||||:||| ||: |||||| |||| ||||||||:|||
Db        345 LSEESQVYTFFSSEFINTINKPVHAALFISWINQVIRDFTTEATQKSTFDKIADISLVVP 404

Qy        605 YIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVI 664
              |:|||||||| || |||:| |||||||||| |||||||||||||||:|||:||||:|
Db        405 YVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELLIPTILVFTIKSFIGSSENKNKII 464
```

```
Qy        665 KAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYN 724
Db        465 KAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTVIEYKYN 524

Qy        725 SYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLRE 784
Db        525 NYTSDERNRLESEYNINNIREELNKKVSLAMENIERFITESSIFYLMKLINEAKVSKLRE 584

Qy        785 YDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFF 844
Db        585 YDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNNSIPFELSSYTNDKILILYFNKLY 644

Qy        845 KRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQND 904
Db        645 KKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNN 704

Qy        905 YIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQ 964
Db        705 DIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQ 763

Qy        965 DNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIH 1024
Db        764 DTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIH 823

Qy       1025 VSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDK 1084
Db        824 VSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNK 883

Qy       1085 EYYLLNVLKPNNFIDRRKDSTLSINNIRS-----TILLANRLYSGIKVKIQRVNNSSTN- 1138
Db        884 RYYLLNLLRTDKSI-TQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRK--NGSTDI 940

Qy       1139 ---DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKI---SSSGNRFNQVVVMNS 1192
Db        941 SNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDS 999

Qy       1193 VGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1251
Db       1000 IGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 1058
```

RESULT 3
AAY93312
ID AAY93312 standard; protein; 1084 AA.
XX
AC AAY93312;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;

XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 57-60; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide dismutase
CC (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC serotype F. The specification describes a composition for delivery of SOD
CC to neuronal cells. The composition comprises SOD linked, by a cleavable
CC linker, to a neuronal cell targeting component (NCTC). This component has
CC a domain that binds to a neuronal cell and a domain that translocates the
CC SOD of the composition into the neuronal cell. After translocation, the
CC linker is cleaved to release the SOD. The composition is useful for
CC treating neuronal diseases caused or augmented by oxidative stress, such
CC as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases
XX
SQ Sequence 1084 AA;

Query Match 46.5%; Score 3040.5; DB 3; Length 1084;
Best Local Similarity 68.2%; Pred. No. 2.3e-159;
Matches 572; Conservative 133; Mismatches 117; Indels 17; Gaps 7;

```
Qy        425 ICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLS 484
Db        250 LCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLNNNYRNNLDEVILDYNSETIPQIS 309

Qy        485 DEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTA 544
Db        310 NQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFFYLHAQKVPEGETNISLTSSIDTA 369

Qy        545 LLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVP 604
Db        370 LSEESQVYTFFSSEFINTINKPVHAALFISWINQVIRDFTTEATQKSTFDKIADISLVVP 429

Qy        605 YIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVI 664
Db        430 YVGLALNIGNEVQKENFKEAPELLGAGILLEFVPELLIPTILVFTIKSFIGSSENKNKII 489

Qy        665 KAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYN 724
Db        490 KAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTVIEYKYN 549

Qy        725 SYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLRE 784
Db        550 NYTSDERNRLESEYNINNIREELNKKVSLAMENIERFITESSIFYLMKLINEAKVSKLRE 609

Qy        785 YDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFF 844
Db        610 YDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNNSIPFELSSYTNDKILILYFNKLY 669

Qy        845 KRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQND 904
Db        670 KKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNN 729

Qy        905 YIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQ 964
Db        730 DIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQ 788

Qy        965 DNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIH 1024
Db        789 DTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIH 848

Qy       1025 VSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDK 1084
Db        849 VSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNK 908
```

```
Qy       1085 EYYLLNVLKPNNFIDRRKDSTLSINNIRS-----TILLANRLYSGIKVKIQRVNNSSTN- 1138
Db        909 RYYLLNLLRTDKSI-TQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRK--NGSTDI 965

Qy       1139 ---DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKI---SSSGNRFNQVVVMNS 1192
Db        966 SNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDS 1024

Qy       1193 VGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1251
Db       1025 IGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 1083
```

RESULT 4
ABG69085
ID ABG69085 standard; protein; 858 AA.
XX
AC ABG69085;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #19.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98555.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 157-159; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 858 AA;

Query Match 41.7%; Score 2728.5; DB 5; Length 858;
Best Local Similarity 61.6%; Pred. No. 2.8e-142;

Matches 527; Conservative 141; Mismatches 156; Indels 31; Gaps 9;

```
Qy          4 INSFNYNDPVNDRTILYIK---PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
Db          5 INSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFDPPAS 64

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
Db         65 LENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGNEHTP 124

Qy        121 DNQFH-IGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNI-SLRNN---YMPSN 175
Db        125 INEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVYDPSN 184

Qy        176 HGFGSIAIVTFSPEYSFRFND------NSMNEFIQDPALTLMHELIHSLHGLYGAKGITT 229
Db        185 DGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGARGVTY 244

Qy        230 KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKV 289
Db        245 KETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATRLSRV 304

Qy        290 QVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKC 347
Db        305 NSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKFKVKC 364

Qy        348 RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKK 407
Db        365 RNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKGLVEK 424

Qy        408 IIRFCKNIVSVKGIR--KSI------------CIEINNGELFFVASENSYNDDNINTPKE 453
Db        425 IVKFCKSVIPRKGTKNCKSVIPRKGTKAPPRLCIRVNNSELFFVASESSYNENDINTPKE 484

Qy        454 IDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHD 513
Db        485 IDDTTNLNNNYRNNLDEVILDYNSQTIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYD 544

Qy        514 VNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFV 573
Db        545 VVDFNVFFYLHAQKVPEGETNISLTSSIDTALLEESK-DIFFSSEFIDTINKPVNAALFI 603

Qy        574 SWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGIL 633
Db        604 DWISKVIRDFTTEATQKSTVDKIADISLIVPYVGLALNIIIEAEKGNFEEAFELLGVGIL 663

Qy        634 LEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKI 693
Db        664 LEFVPELTIPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRI 723

Qy        694 NTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSI 753
Db        724 NTQFNKRKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESEYNINNIEEELNKKVSL 783

Qy        754 AMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNS 813
Db        784 AMKNIERFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLLNYILDHRSILGEQTNELSD 843

Qy        814 MVTDTLNNSIPFKLS 828
Db        844 LVTSTLNSSIPFELS 858
```

RESULT 5
AAY77137
ID AAY77137 standard; protein; 449 AA.
XX
AC AAY77137;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype E (BoNTE) C-terminal fragment.
XX

KW Botulinum neurotoxin; heavy chain; BoNT; serotype E; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87215.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 26; Page 43-44; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 449 AA;

Query Match 36.4%; Score 2384; DB 3; Length 449;
Best Local Similarity 99.3%; Pred. No. 1.3e-123;
Matches 446; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy        804 LGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 863
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 60

Qy        864 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRI 923
              ||||||||||||||||||||||||||||||:|:|||||||||||||||||||||||||||
Db         61 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLTELNISQNDYIIYDNKYKNFSISFWVRI 120

Qy        924 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGI 983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGI 180

Qy        984 SDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIG 1043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIG 240

Qy       1044 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKD 1103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKD 300

Qy       1104 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLY 1163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLY 360

Qy       1164 ADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAST 1223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAST 420

Qy       1224 WYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||
Db        421 WYYTHMRDHTNSNGCFWNFISEEHGWQEK 449
```

RESULT 6
AAB04094
ID AAB04094 standard; protein; 449 AA.
XX
AC AAB04094;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54488.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 7b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.

```
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 449 AA;

  Query Match             36.4%;  Score 2384;  DB 4;  Length 449;
  Best Local Similarity   99.3%;  Pred. No. 1.3e-123;
  Matches  446;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        804 LGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 863
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 60

Qy        864 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRI 923
              ||||||||||||||||||||||||||||||:|:|||||||||||||||||||||||||||
Db         61 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLTELNISQNDYIIYDNKYKNFSISFWVRI 120

Qy        924 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGI 983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGI 180

Qy        984 SDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIG 1043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIG 240

Qy       1044 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKD 1103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKD 300

Qy       1104 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLY 1163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 STLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLY 360

Qy       1164 ADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAST 1223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAST 420

Qy       1224 WYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||
Db        421 WYYTHMRDHTNSNGCFWNFISEEHGWQEK 449


RESULT 7
AAR95010
ID   AAR95010 standard; protein; 1296 AA.
XX
AC   AAR95010;
XX
DT   09-JUL-1996  (first entry)
XX
DE   C. botulinum type A neurotoxin.
XX
KW   Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.
XX
OS   Clostridium botulinum.
XX
PN   WO9612802-A1.
XX
PD   02-MAY-1996.
XX
PF   23-OCT-1995;   95WO-US013737.
XX
PR   24-OCT-1994;   94US-00329154.
PR   16-MAR-1995;   95US-00405496.
PR   14-APR-1995;   95US-00422711.
PR   07-JUN-1995;   95US-00480604.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Padhye NV,  Kink JA,  Thalley BS,  Stafford DC;
PI   Firca JR;
XX
DR   WPI; 1996-230603/23.
DR   N-PSDB; AAT29244.
XX
PT   Fusion proteins comprising non-toxin protein and part of toxin - useful
PT   to form anti-toxins against Clostridium botulinum type A, and C.
PT   difficile type toxins, and to treat C. difficile intoxication, partic.
PT   diarrhoea.
XX
PS   Claim 4; Page 344-350; 434pp; English.
XX
CC   Clostridium botulinum type A neurotoxin (AAR95010) is processed to form a
CC   dimer composed of a light and a heavy chain. It is the product of the
CC   type A neurotoxin gene (AAT29244). The 50 kDa C-terminal portion of the
CC   heavy chain, or C fragment (see also AAR95008), was produced using a
CC   synthetic gene (AAT29245) having codon usage altered to improve
CC   expression in Escherichia coli. Fusion proteins of the type A toxin or C
CC   fragment, with e.g. maltose binding protein or polyhistidine affinity tag
CC   (see also AAR95008), are used to generate neutralising antitoxins and in
CC   vaccine compsns
XX
SQ   Sequence 1296 AA;

  Query Match             35.7%;  Score 2336.5;  DB 2;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 2.1e-120;
  Matches  534;  Conservative  247;  Mismatches  418;  Indels  135;  Gaps   43;

Qy          7 FNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |:|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::          :|  |||:|| |||||: | ||    ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              | |   ||    : |  :  ||  |||| |   | | | |:         :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:||  |              :||::|| :|||  ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |        : : |  | || :: |  :| | |:
Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |     :|  :  ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
              ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696
```

```
Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: | |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNI 873

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYII 907
               ::|:||:||:::  :| | | | |||   |   | :|||  ::| : |:: :   : |:
Db        874 INTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIV 933

Qy        908 YDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNA 967
              |:: |:||| |||:||| | | | ::|||||||||| :|||||||||: ||||||||
Db        934 YNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIWTLQDTQ 991

Qy        968 GINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD 1027
              | |:: | |     ||||||:||||||||:|| :||:|||| ||||| | |||||| |:
Db        992 EIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASN 1051

Qy       1028 NILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEY 1086
              ||:||:  |  | ||| |:|||:|||||:| ||: || |: |: |||||||:|| ||| |
Db       1052 NIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPY 1111

Qy       1087 YLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYSGIKVKIQ 1130
              |:||:  || ::|        :||:                : | | : || | |  |:
Db       1112 YMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIK 1163

Qy       1131 RVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSSGNRFNQV 1187
              :  |   ||:|| ||:|||| |  |   : |  : :    ||    ::|   ||  :||
Db       1164 KY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDVGN-LSQV 1220

Qy       1188 VVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDHTNSNGC 1238
              ||| |     : | | || ::||||:|| :||        :||| ||   :   : : ||
Db       1221 VVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSSRTLGC 1280

Qy       1239 FWNFISEEHGWQEK 1252
               | ||  : || |:
Db       1281 SWEFIPVDDGWGER 1294
```

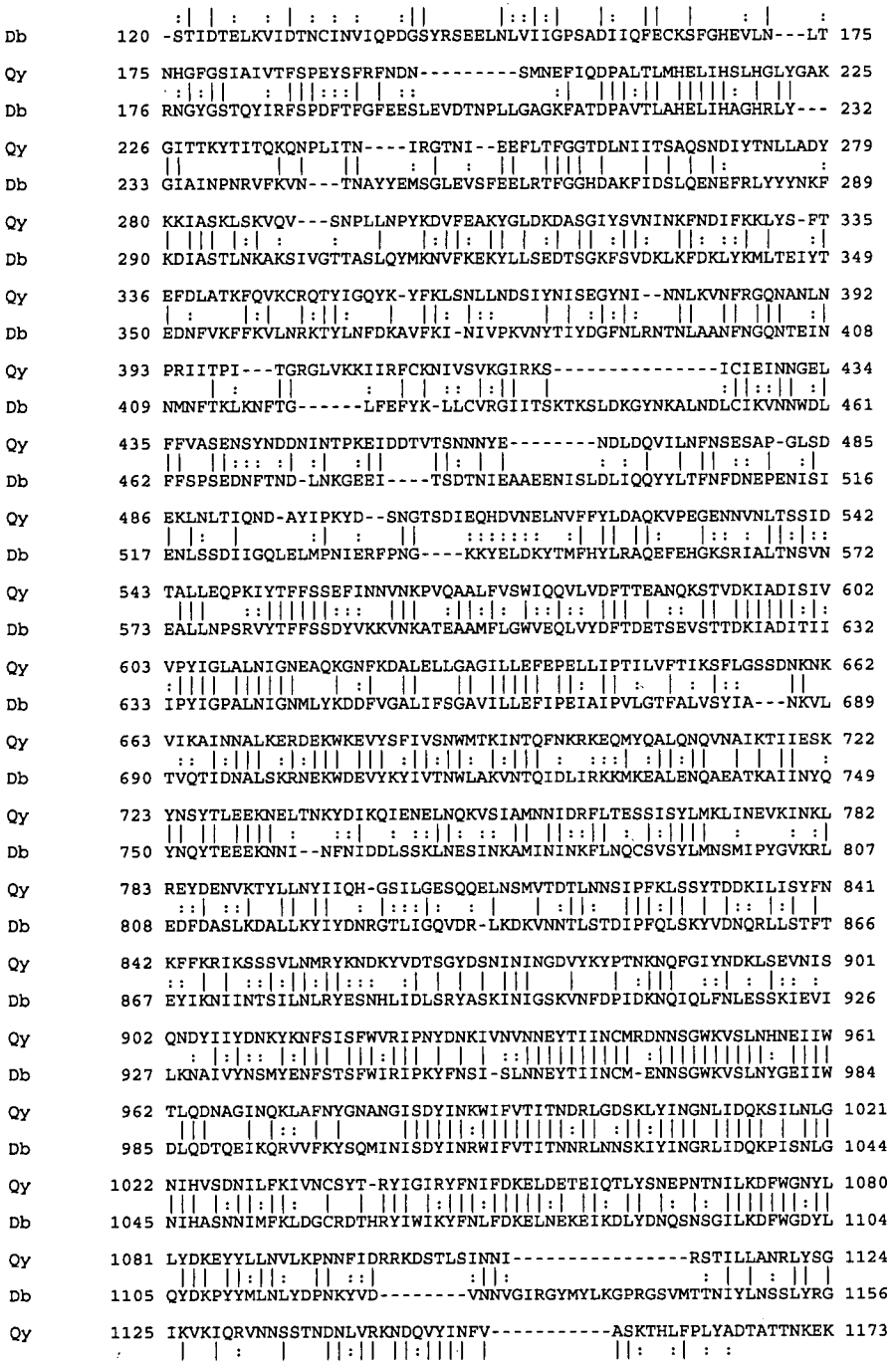
RESULT 8
AAU99339
ID AAU99339 standard; protein; 1295 AA.
XX
AC AAU99339;
XX
DT 07-OCT-2002 (first entry)
XX
DE Clostridium botulinum botulism A toxin protein.
XX
KW Botulism A toxin; neuroprotective; muscular; BoTOX; tetanus toxin; TeTOX;
KW protease cleavage; toxicity; neurotoxin; poison; neural cell; toxin;
KW post-translational modification; light chain; heavy chain;
KW spastic condition; stabismus; bephlarospasm; hemifacial spasm;
KW brain injury; spinal cord injury; stroke; multiple sclerosis;
KW cerebral palsy; BoNT/A.
XX
OS Clostridium botulinum.
XX
FH Key Location/Qualifiers
FT Region 1. .437
FT /note= "Light chain"
FT Active-site 222
FT /note= "Forms an active site with Glu223, His226, Glu261,
FT Tyr365"
FT Active-site 223
FT /note= "Forms an active site with His222, His226, Glu261,
FT Tyr365"
FT Active-site 226
FT /note= "Forms an active site with His222, Glu223, Glu261,
FT Tyr365"
FT Active-site 261
FT /note= "Forms an active site with His222, Glu223, His226,
FT Tyr365"
FT Active-site 365
FT /note= "Forms an active site with His222, Glu223, His226,
FT Glu261"
FT Region 437. .448
FT /note= "Linker region"
FT Region 449. .1295
FT /note= "Heavy chain"
FT Misc-difference 1161
FT /label= OTHER
FT /note= "OTHER = O, which is not defined in the
FT specification"
XX
PN WO200244199-A2.
XX
PD 06-JUN-2002.
XX
PF 27-NOV-2001; 2001WO-US045059.
XX
PR 29-NOV-2000; 2000US-00726949.
XX
PA (ALLR) ALLERGAN SALES INC.
XX
PI Lin W, Aoki KR, Steward LE;
XX
DR WPI; 2002-557531/59.
XX
PT Novel modified botulism toxin or tetanus toxin comprising a protease
PT cleavage site, is useful for treating conditions benefited by neurotoxin
PT activity.
XX
PS Disclosure; Fig 1; 19pp; English.
XX
CC The invention discloses modified botulism toxin (BoTOX) or tetanus toxin
CC (TeTOX) which contain new protease cleavage sites. These sites are
CC created in regions where the cleavage site is specifically susceptible to
CC cleavage by a protease that is present in effective levels only in a
CC tissue where toxicity is undesirable and where the cleavage site itself
CC does not inactivate the toxin. BoTOX and TeTOX are neurotoxins which are
CC highly potent and specific poisons of neural cells. The toxins are
CC synthesised in vivo as single chains, which are not toxic, but become
CC active when nicked, in a post-translational modification, to form
CC separate light and heavy chains which are linked through a disulphide
CC bond. The modified toxins are useful to treat conditions benefited by
CC neurotoxin activity (e.g. spastic conditions, including stabismus,
CC bephlarospasm and hemifacial spasm, brain injury, spinal cord injury,
CC stroke, multiple sclerosis and cerebral palsy) by administering the
CC toxins for their localised production. The advantage of the toxins are
CC that they are deactivated in tissues where toxic activity is undesirable
CC and activated at desired targets. The sequence presented is the
CC Clostridium botulinum botulism A toxin (BoNT/A) protein
XX
SQ Sequence 1295 AA;

Query Match 35.5%; Score 2324; DB 5; Length 1295;
Best Local Similarity 39.6%; Pred. No. 1e-119;
Matches 534; Conservative 248; Mismatches 414; Indels 152; Gaps 44;

```
Qy          2 PKIN-SFNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFH 56
              | :|  ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :
Db          1 PFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLN 59

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
               ||   |    ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        117 DNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPS 174
```



Db 120 -STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LT 175

Qy 175 NHGFGSIAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAK 225

Db 176 RNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY--- 232

Qy 226 GITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADY 279

Db 233 GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKF 289

Qy 280 KKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 335

Db 290 KDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT 349

Qy 336 EFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 392

Db 350 EDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEIN 408

Qy 393 PRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS----------ICIEINNGEL 434

Db 409 NMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDL 461

Qy 435 FFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSD 485

Db 462 FFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISI 516

Qy 486 EKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSID 542

Db 517 ENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVN 572

Qy 543 TALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIV 602

Db 573 EALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITII 632

Qy 603 VPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNK 662

Db 633 IPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVL 689

Qy 663 VIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESK 722

Db 690 TVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ 749

Qy 723 YNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKL 782

Db 750 YNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRL 807

Qy 783 REYDENVKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFN 841

Db 808 EDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFT 866

Qy 842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901

Db 867 EYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVI 926

Qy 902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961

Db 927 LKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIW 984

Qy 962 TLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLG 1021

Db 985 DLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG 1044

Qy 1022 NIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Db 1045 NIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYL 1104

Qy 1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNI----------RSTILLANRLYSG 1124

Db 1105 QYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy 1125 IKVKIQRVNNSSTNDNLVRKNDQVYINFV----------ASKTHLFPLYADTATTNKEK 1173

Db 1157 TKF-IXKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQVVVFKILS--------- 1206

Qy 1174 TIKISSSGNRFNQVVVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTW 1224

Db 1207 ALEIPDVGN-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNW 1265

Qy 1225 YYTHMRDHTNSNGCFWNFISEEHGWQEK 1252

Db 1266 YNRQIERSSRTLGCSWEFIPVDDGWGER 1293

RESULT 9
AAW68396
ID AAW68396 standard; protein; 452 AA.
XX
AC AAW68396;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain NCTC 11219.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30585.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 41; Page 327-329; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC express BotE soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 452 AA;

Query Match 34.7%; Score 2267; DB 2; Length 452;
Best Local Similarity 99.8%; Pred. No. 3.9e-117;
Matches 425; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 827 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 886

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 86

Qy        887 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 946
                |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 QFGIDNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 146

Qy        947 NSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 1006
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 NSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 206

Qy       1007 INGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE 1066
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 INGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE 266

Qy       1067 PNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIK 1126
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 PNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIK 326

Qy       1127 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQ 1186
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQ 386

Qy       1187 VVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEE 1246
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 VVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEE 446

Qy       1247 HGWQEK 1252
                ||||||
Db        447 HGWQEK 452
```

RESULT 10
AAB04095
ID AAB04095 standard; protein; 419 AA.
XX
AC AAB04095;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54489.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Fig 8; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 419 AA;

Query Match 34.1%; Score 2230; DB 4; Length 419;
Best Local Similarity 99.5%; Pred. No. 3.9e-115;
Matches 416; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 61

Qy        895 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 954
                |:|:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 LTELNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 121

Qy        955 NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQ 1014
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQ 181

Qy       1015 KSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKD 1074
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 KSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKD 241

Qy       1075 FWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNN 1134
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 FWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNN 301

Qy       1135 SSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 1194
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 SSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG 361

Qy       1195 NNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 NNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 419
```

RESULT 11
ABG69071
ID ABG69071 standard; protein; 422 AA.
XX
AC ABG69071;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #5.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX

PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98541.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 125-126; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 422 AA;

Query Match 33.8%; Score 2212; DB 5; Length 422;
Best Local Similarity 100.0%; Pred. No. 3.9e-114;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IR 422
              ||
Db        421 IR 422
```

RESULT 12
AAW68395
ID AAW68395 standard; protein; 451 AA.
XX
AC AAW68395;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain Belgua.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30584.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 41; Page 324-325; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC express BotE soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 451 AA;

Query Match 33.8%; Score 2210.5; DB 2; Length 451;
Best Local Similarity 97.7%; Pred. No. 5.1e-114;
Matches 416; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

```
Qy        827 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 LSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKN 86

Qy        887 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 QFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDN 146

Qy        947 NSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 1006
```

```
                    ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        147 NSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLY 206

Qy       1007 INGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE 1066
                    |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        207 INGNLIDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNE 266

Qy       1067 PNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIK 1126
                    || ||||||||||||||||||||||||||||:|| ||||||||||||||||||||||||
Db        267 PNANILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIK 326

Qy       1127 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQ 1186
                    |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        327 VKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQ 386

Qy       1187 VVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEE 1246
                    |||||||| ||||||||||||||||||||||||||||||||:||||| ||||||||
Db        387 VVVMNSVG-NCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEE 445

Qy       1247 HGWQEK 1252
                    ||||||
Db        446 HGWQEK 451
```

RESULT 13
ABG69082
ID ABG69082 standard; protein; 427 AA.
XX
AC ABG69082;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #16.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98552.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 150-151; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 427 AA;

```
  Query Match             33.8%;  Score 2210;  DB 5;  Length 427;
  Best Local Similarity   99.8%;  Pred. No. 5.1e-114;
  Matches  422;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFXRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRK 423
              |||
Db        421 IRK 423
```

RESULT 14
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;

```
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30579.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 291-296; 428pp; English.
XX
CC   This is the amino acid sequence of the type B toxin of Clostridium
CC   botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC   B toxin has been expressed as a histidine-tagged protein in Escherichia
CC   coli. The invention relates to C. botulinum recombinant toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of the recombinant proteins are E. coli, insect cells and
CC   yeast cells. The recombinant proteins are used as immunogens for the
CC   production of vaccines and antitoxins that are useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin.
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1291 AA;

  Query Match             33.5%;  Score 2194;  DB 2;  Length 1291;
  Best Local Similarity   37.4%;  Pred. No. 1.5e-112;
  Matches  491;  Conservative  271;  Mismatches  468;  Indels   82;  Gaps   32;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||||:::  |: ::|    |   :||:| |   |||||||   | |:||:
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : : | |   ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   |  :|:   ||   |  :   ::    :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   ||:   |:          :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :       ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |      |  :| ||: |: ||   :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|       |: |||:: || | ||:||:  :::  :||||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:   |    |: ||   |||     |||:::| :|||:| :|||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  |||  ::::||: :  |   |  |         :  :| |:       |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
               :|  :| || :|      :::|||| | |||   |:|:||| ::|    || |:| |
Db        530 IFTDENTIFQYLYSQTFLLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::||||||||||:|||  ||||::| |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
```

```
              ||||| |||||| :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKV 751
              :||||   || ||:||  |  |:: ||: :|| |: :||: :    |   | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQEL 811
              : |::||: |:   |:|||| :  : : || ::|  :|  ||||| ::   |  | :
Db        765 NQAIHNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
               | |   |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |     :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINK 989
              ::||||||||:|||||:|:  | ||||| |  |  : : | |     ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | ||||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS---- 1104
              ||: || :: |:  |  :  :  |||||| |:|:||||: |    |::|  :|||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFP 1161
                | |  |  |  :    || | |  :| :|| | ||::||| | :|::|  :   :
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWR 1179

Qy       1162 LYADTATTNKEKTI---KISSSGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLG 1213
              :|       :|: :    || |  :| : :       :| : || :  : : |||:|
Db       1180 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYHEQPTYSCQLLFKKDEESTDEIGLIG 1239

Qy       1214 FK------------ADTVVASTWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251
                            |    | ||   :  : :    || | || :: || |
Db       1240 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 15
AAB61169
ID   AAB61169 standard; protein; 1315 AA.
XX
AC   AAB61169;
XX
DT   02-APR-2001  (first entry)
XX
DE   Clostridium tetani TeNT.
XX
KW   Clostridium tetani; TeNT; tetanus toxin; antibacterial; vaccine;
KW   TeNT fragment C; infection.
XX
OS   Clostridium tetani.
XX
PN   WO200100839-A1.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-GB002428.
XX
PR   25-JUN-1999;   99GB-00014861.
XX
PA   (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI   Fairweather NF,  Sinha K;
XX
DR   WPI; 2001-123014/13.
XX
```

PT New polypeptide, useful for treating infections of Clostridium tetani,
PT comprises tetanus toxin fragment with a mutation in a loop region,.
XX
PS Disclosure; Page 39; 43pp; English.
XX
CC The present sequence is given in a specification relating to a novel
CC polypeptide comprising tetanus toxin (TeNT) fragment C or its immunogenic
CC fragment, containing a mutation in a loop region. The mutation results in
CC a reduction in the binding of TeNT fragment C or its immunogenic fragment
CC to gangliosides and primary motoneurones, and/or a reduction in the
CC ability of TeNT fragment C or its immunogenic fragment to undergo
CC retrograde transport. The polypeptide is useful for treating, preventing
CC and reducing the susceptibility to Clostridium tetani infection in a
CC human or animal, and also for producing antibodies which recognise groups
CC within TeNT polypeptides. Antibody produced against the polypeptide is
CC also useful for treating Clostridium tetani infection
XX
SQ Sequence 1315 AA;

Query Match 32.9%; Score 2151.5; DB 4; Length 1315;
Best Local Similarity 35.8%; Pred. No. 3.4e-110;
Matches 490; Conservative 260; Mismatches 444; Indels 173; Gaps 37;

```
Qy          1 MP-KINSFNYNDPVNDRTILYIKPGGCQE---FYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:| |:|||| ||: ::|  |:    :||:| |   |||:|||   || |:||:
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
              ||:||  | | |||||||:| :|||||: : |:|||| ||::|  ||:::  | |||||
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        117 DNTPDNQFHIGDASAVEIKFSNGSQD--------ILLPNVIIMGAEPDLFETNSSNISLR 168
                :  ::|   | ::  : |:   ||         :| |:|| |  | | :     | ||
Db        121 SYSLLDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLR 177

Qy        169 ---NNYMPSNHGFGSIAIVTFSPEYSFRFNDN----------SMNEFIQDPALTLMHELI 215
                 || |   ||||| : | |||   | ||            ::: ||||| ||||||
Db        178 VDNKNYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELI 236

Qy        216 HSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275
              | |||||| : :::   |  ||   : :    : ||  |||| | |:|:    ||:|
Db        237 HVLHGLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKT 295

Qy        276 LADYKKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-L 331
              | ||| ||:|||:|   |      :: || ::: ||  |||::| | || :||  ::   :
Db        296 LNDYKAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIM 355

Qy        332 YSFTEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388
              | ||| :|  || :| | :|    :   |: |||:|:||| :||:||   :||  ::|||
Db        356 YGFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQN 415

Qy        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKS-------------ICIEINNGEL 434
                :|      : | ||| |:|  || |:    ||::                :||:| | :|
Db        416 MRVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDL 475

Qy        435 FFVASENSYNDDNINTPKEIDDTVTSNN-----NYENDLDQVILNFNSESAPGLSDEKLN 489
              |:| :||:::::         |: |: |       |:   ||::|:::| :|   | :::
Db        476 TFIAEKNSFSEEPFQ-----DEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTT 530

Qy        490 LTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQP 549
                :   | |:| ||  | || |:::: :: || ||| |     : :|:|:| ||:
Db        531 PVTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLGRITMTNSVDDALINST 590

Qy        550 KIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLA 609
              |||::| |  |: ||:  |  ||: |:: :: ||| |::||:|:|:|||:|:| :||||| |
Db        591 KIYSYFPS-VISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPA 649

Qy        610 LNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINN 669
              |||  :  :|||  |||  |  :|||: ||: :| |   :|     ||  | |:|| |:|
Db        650 LNIVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAE---SSTQKEKIIKTIDN 706

Qy        670 ALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLE 729
               |::| ||| |||  : : |:  :|||| ||  |||::|: ||:||| ||: :|  |:

Db        707 FLEKRYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGP 766

Qy        730 EKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLM-KLINEVKINKLREYDEN 788
              :| :: :  :|  ::|:| :| : || ||: |: ||| |:|: ::||| |  :| |:|
Db        767 DKEQIAD--EINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAK-KQLLEFDTQ 823

Qy        789 VKTYLLNYIIQHGSILGESQ-QELNSMVTDTLNNSIPFKLSSYTD---------DKILIS 838
              |  |: ||  :  :| :: ::| | :    :  |||  |  |          | ||
Db        824 SKNILMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL-- 881

Qy        839 YFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSE 897
                    | |::||:   ||   | ||::|::    |    |  |     : |:: ||
Db        882 ---------KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSE 932

Qy        898 VNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNV--NNEYTIINCMRDNN----SGWK 951
              | : :  | |:: : ||::|||:|:|     :     |||:||: |: ::    |||
Db        933 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 992

Qy        952 VSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYI-NKWIFVTITNDRLGDSKLYINGN 1010
              |||  | :||||:|:||  ::: |     :  : |: |||:|:||||||  : |||||
Db        993 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 1051

Qy       1011 LIDQKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069
              |:    |  || |   :||  |:  | :  :|: |  | || | |:  ||: ||::  :
Db       1052 LMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSI 1111

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLAN---------- 1119
                |:||||| | || ||||: |   :      ||  : : ||   : | |
Db       1112 TFLRDFWGNPLRYDTEYYLIPVASSS------KD--VGLKNITDYMYLTNAPSYTNGKLN 1163

Qy       1120 ----RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLFPLYADTATTNKEK 1173
                   |||:|:|  |:|   ::  |: |:  |  ::|:::                 | |
Db       1164 IYYRRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSY---------------NNNEH 1207

Qy       1174 TIKISSSGNRFNQVVVMNSVGNNC---------------------TMNFKNNNGNNIGL 1211
                :     || || :  :  || |                          :  ::   ::||
Db       1208 IVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGL 1267

Qy       1212 LGF---------KADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGW 1249
              :|           | ::|| ||: |::|     || | |:  : ||
Db       1268 VGTHNGQIGNDPNRDILIASNWYFNHLKD--KILGCDWYFVPTDEGW 1312
```

Search completed: September 10, 2004, 03:33:34
Job time : 85.5862 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 26.574 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 6541
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCPWNFISEEHGWQEK 1252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2336.5	35.7	1296	1	US-08-480-604A-28	Sequence 28, Appl
2	2336.5	35.7	1296	2	US-08-405-496A-28	Sequence 28, Appl
3	2336.5	35.7	1296	3	US-08-915-136-28	Sequence 28, Appl
4	2336.5	35.7	1296	4	US-09-084-517-28	Sequence 28, Appl
5	2151.5	32.9	1315	4	US-08-913-880C-1	Sequence 1, Appli
6	2141.5	32.7	1169	4	US-09-255-829-20	Sequence 20, Appl
7	1526.5	23.3	858	4	US-09-255-829-22	Sequence 22, Appl
8	1526.5	23.3	858	4	US-09-255-829-29	Sequence 29, Appl
9	1359.5	20.8	1013	4	US-09-255-829-18	Sequence 18, Appl
10	1353	20.7	907	4	US-09-255-829-16	Sequence 16, Appl
11	1348	20.6	871	4	US-09-255-829-8	Sequence 8, Appli
12	1340.5	20.5	871	4	US-09-255-829-2	Sequence 2, Appli
13	1340.5	20.5	871	4	US-09-255-829-26	Sequence 26, Appl
14	1340.5	20.5	875	4	US-09-255-829-10	Sequence 10, Appl
15	1338.5	20.5	873	4	US-09-255-829-6	Sequence 6, Appli
16	1338.5	20.5	894	4	US-09-255-829-4	Sequence 4, Appli
17	1338	20.5	878	4	US-09-255-829-12	Sequence 12, Appl
18	1338	20.5	953	4	US-09-255-829-14	Sequence 14, Appl
19	1308	20.0	858	4	US-08-913-880C-16	Sequence 16, Appl
20	1308	20.0	860	4	US-08-913-880C-15	Sequence 15, Appl
21	1308	20.0	862	4	US-08-913-880C-14	Sequence 14, Appl
22	1308	20.0	865	4	US-08-913-880C-13	Sequence 13, Appl
23	1308	20.0	866	4	US-08-913-880C-12	Sequence 12, Appl
24	1308	20.0	874	4	US-08-913-880C-11	Sequence 11, Appl
25	1308	20.0	875	4	US-08-913-880C-10	Sequence 10, Appl
26	1307	20.0	853	4	US-08-913-880C-17	Sequence 17, Appl
27	1020	15.6	438	1	US-08-480-604A-23	Sequence 23, Appl
28	1020	15.6	438	2	US-08-405-496A-23	Sequence 23, Appl
29	1020	15.6	438	3	US-08-915-136-23	Sequence 23, Appl
30	1020	15.6	438	4	US-09-084-517-23	Sequence 23, Appl
31	1020	15.6	462	1	US-08-480-604A-26	Sequence 26, Appl
32	1020	15.6	462	2	US-08-405-496A-26	Sequence 26, Appl
33	1020	15.6	462	3	US-08-915-136-26	Sequence 26, Appl
34	1020	15.6	462	4	US-09-084-517-26	Sequence 26, Appl
35	782	12.0	548	4	US-09-255-829-24	Sequence 24, Appl
36	580	8.9	618	1	US-08-668-381A-5	Sequence 5, Appli
37	579.5	8.9	452	1	US-07-618-312A-2	Sequence 2, Appli
38	579.5	8.9	452	1	US-08-280-228-2	Sequence 2, Appli
39	575.5	8.8	452	1	US-07-618-312A-4	Sequence 4, Appli
40	575.5	8.8	452	1	US-08-280-228-4	Sequence 4, Appli
41	546.5	8.4	452	1	US-08-110-786A-8	Sequence 8, Appli
42	368.5	5.6	207	3	US-08-446-114A-21	Sequence 21, Appl
43	309	4.7	2184	4	US-09-417-485D-6	Sequence 6, Appli
44	264.5	4.0	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
45	259.5	4.0	2710	1	US-08-480-604A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match           35.7%;  Score 2336.5;  DB 1;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 5e-144;
  Matches  534;  Conservative  247;  Mismatches  418;  Indels  135;  Gaps   43;

Qy          7 FNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :||  |
Db          8 FNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| | |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : :   :||       |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
                : |||:::| | ::           :|  |||:|| |||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
                | |   ||    : |  :  ||  |||| |   | | | |:          :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:||  |              :||::|| :|||  ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |         : :  |  | || :: |  :| | |:
Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |     :|  :   ||    ::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
               ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||    | :|  ||   || ||||| ||: || :  | : |::   ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||   | | ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847
                :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: | |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNI 873

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYII 907
              ::|:||:||:::  :| | | | |||   |   | :|||  ::| : |:: :   : |:
Db        874 INTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIV 933

Qy        908 YDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNA 967
              |:: |:||| |||:||| | | | ::|||||||||| :|||||||||: |||||||
Db        934 YNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIWTLQDTQ 991

Qy        968 GINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD 1027
              | |:: | |     ||||||:||||||||:|| :||:|||| ||||| | |||||| |:
Db        992 EIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASN 1051

Qy       1028 NILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLLYDKEY 1086
              ||:||:  |  | ||| |:|||:|||||:| ||: || |: |: |||||||:|| ||| |
Db       1052 NIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPY 1111

Qy       1087 YLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYSGIKVKIQ 1130
              |:||:  || ::|        :||:                  : | | : || | |  |:
Db       1112 YMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIK 1163

Qy       1131 RVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSSGNRFNQV 1187
              :  |  ||:|| ||:|||| |  |  : |  : :   ||   ::|   ||  :||
Db       1164 KY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDVGN-LSQV 1220

Qy       1188 VVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDHTNSNGC 1238
              ||| |     : | | || ::||||:|| :||        :||| ||   :  : : ||
Db       1221 VVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSSRTLGC 1280

Qy       1239 FWNFISEEHGWQEK 1252
                | ||  : || |:
Db       1281 SWEFIPVDDGWGER 1294


RESULT 2
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             35.7%;  Score 2336.5;  DB 2;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 5e-144;
  Matches  534;  Conservative  247;  Mismatches  418;  Indels  135;  Gaps   43;

Qy          7 FNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
                : |||:::| | ::            :|  |||:|| ||||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
                | |   ||    : |  :  ||  |||| |   | | | |:          :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
               |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:||  |               :||::|| :|||  ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |         : :  |  | || :: |  :| | |:
Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
               ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:|:|: |:||| :  :::| :||:||  | |  ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847
              :|  || ||  : |:::|:  : |    | :||:  |||:|| | |:: |:| | :: | |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNI 873

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYII 907
              ::|:||:||::  :| | | | |||    |   | :|||  ::| : |:: :   : |:
Db        874 INTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIV 933

Qy        908 YDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNA 967
              |:: |:||| |||:||| | | | ::|||||||||| :|||||||||: ||||||||
Db        934 YNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIWTLQDTQ 991

Qy        968 GINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD 1027
              | |:: | |      ||||||:|||||||:|| :||:|||| ||||| | |||||| |:
Db        992 EIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASN 1051

Qy       1028 NILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEY 1086
              ||:||:  |  | ||| |:|||:|||||:| ||: || |: |: |:| ||||||:|| ||| |
Db       1052 NIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPY 1111

Qy       1087 YLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYSGIKVKIQ 1130
              |:||:  || ::|        :||:                 : | | : || | |  |:
Db       1112 YMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIK 1163

Qy       1131 RVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSSGNRFNQV 1187
              :  |   ||:|| ||:|||| |  |   : |  : :    ||    ::|   ||  :||
Db       1164 KY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDVGN-LSQV 1220

Qy       1188 VVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDHTNSNGC 1238
              ||| |      : | | || ::||||:|| :||         :||| ||   :  : : ||
Db       1221 VVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSSRTLGC 1280

Qy       1239 FWNFISEEHGWQEK 1252
              | ||  : || |:
Db       1281 SWEFIPVDDGWGER 1294


RESULT 3
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
```

```
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             35.7%;  Score 2336.5;  DB 3;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 5e-144;
  Matches  534;  Conservative  247;  Mismatches  418;  Indels  135;  Gaps   43;

Qy          7 FNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                |||| || :| ||| :|| |||:| || :   | :||  : :  |: |  :| | |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |:|:|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::          :|  |||:|| |||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              | |   ||    : |  :  ||  |||| |   | | | |:         :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |  |:||: || | :| || :|:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGELFFVASE 440
              :   ||       :  | | :: |:||  |              :||::|| :|||  ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |        : :  |  | || :: |  :| | |:
Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
                ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: | |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNI 873

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYII 907
              ::|:||:||:::  :| | | | |||   |   | :|||  ::| : |:: :   : |:
Db        874 INTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIV 933

Qy        908 YDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNA 967
              |:: |:||| |||:||| | | | ::|||||||||| :||||||||||: |||||||
Db        934 YNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIWTLQDTQ 991

Qy        968 GINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD 1027
              | |:: | |      ||||||:||||||||:|| :||:|||| ||||| | |||||| |:
Db        992 EIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASN 1051

Qy       1028 NILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEY 1086
              ||:||:  |  | ||| |:|||:|||||:| ||: || |: |: |||||||:|| ||| |
Db       1052 NIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPY 1111

Qy       1087 YLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYSGIKVKIQ 1130
              |:||:  || ::|          :||:                : | | : || | |  |:
Db       1112 YMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIK 1163

Qy       1131 RVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSSGNRFNQV 1187
              :   |   ||:|| ||:|||| |  |   : |  : :    ||    ::|   ||  :||
Db       1164 KY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDVGN-LSQV 1220

Qy       1188 VVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDHTNSNGC 1238
              ||| |     : | | || ::||||:|| :||         :||| ||   :   : : ||
Db       1221 VVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSSRTLGC 1280

Qy       1239 FWNFISEEHGWQEK 1252
              | ||  : || |:
Db       1281 SWEFIPVDDGWGER 1294


RESULT 4
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match           35.7%;  Score 2336.5;  DB 4;  Length 1296;
  Best Local Similarity  40.0%;  Pred. No. 5e-144;
  Matches  534;  Conservative  247;  Mismatches  418;  Indels  135;  Gaps   43;

Qy          7 FNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::           :|  |||:|| ||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              | |   ||    : |  :  ||  |||| |   | | | |:         :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
               |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:||  |                :||::|| :|||  ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |        : : |  | || :: |  :| | |:

Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
              ::||||||:::  |||  :||:|: |::|:: ||| | :: || |||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: | |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNI 873

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYII 907
               ::|:||:||::  :| | | | |||   |   | :|||  ::| : |:: :   : |:
Db        874 INTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIV 933

Qy        908 YDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNA 967
              |:: |:||| |||:||| | | | ::|||||||||| :|||||||||: ||||||||
Db        934 YNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIWTLQDTQ 991

Qy        968 GINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD 1027
              | |:: | |      ||||||:|||||||:|| :||:|||| ||||| | |||||| |:
Db        992 EIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASN 1051

Qy       1028 NILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEY 1086
              ||:||:  |  | ||| |:|||:|||||:| ||: || |: |: |||||||:|| ||| |
Db       1052 NIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPY 1111

Qy       1087 YLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYSGIKVKIQ 1130
              |:||:  || ::|         :||:                : | | : || | |  |:
Db       1112 YMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIK 1163

Qy       1131 RVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSSGNRFNQV 1187
              :   |   ||:|| ||:|||| |  |   : |  : :    ||    ::|   ||  :||
Db       1164 KY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDVGN-LSQV 1220

Qy       1188 VVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDHTNSNGC 1238
              ||| |     : | | || ::||||:|| :||        :||| ||   :  : : ||
Db       1221 VVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSSRTLGC 1280

Qy       1239 FWNFISEEHGWQEK 1252
              | ||  : || |:
Db       1281 SWEFIPVDDGWGER 1294


RESULT 5
US-08-913-880C-1
; Sequence 1, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
;  APPLICANT: MATSUDA, Morihiro
;  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
;  TITLE OF INVENTION:  VACCINE
;  FILE REFERENCE: 216-380P
;  CURRENT APPLICATION NUMBER: US/08/913,880C
;  CURRENT FILING DATE:  1997-11-24
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
;   LENGTH: 1315
```

```
;   TYPE: PRT
;   ORGANISM: Clostridium tetani
US-08-913-880C-1

  Query Match             32.9%;  Score 2151.5;  DB 4;  Length 1315;
  Best Local Similarity   35.8%;  Pred. No. 6e-132;
  Matches  490;  Conservative  260;  Mismatches  444;  Indels  173;  Gaps   37;

Qy          1 MP-KINSFNYNDPVNDRTILYIKPGGCQE---FYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:| |:|||||: ||: ::|   |:    :||:| |   |||:|||   || |:||:
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
              ||:||  | | |||||||::| :|||||: : |:|||| ||::|  ||:::  | |||||
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        117 DNTPDNQFHIGDASAVEIKFSNGSQD--------ILLPNVIIMGAEPDLFETNSSNISLR 168
                :  ::|   | ::  : |:   ||          :| |:|| |  | | :     | ||
Db        121 SYSLLDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLR 177

Qy        169 ---NNYMPSNHGFGSIAIVTFSPEYSFRFNDN----------SMNEFIQDPALTLMHELI 215
                 || |   |||||  : | |||   | ||           ::: |||||  ||||||
Db        178 VDNKNYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELI 236

Qy        216 HSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275
              | ||:|||| : :::   |  ||   : :   : ||  |||| | |:|:     ||:|
Db        237 HVLHGLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKT 295

Qy        276 LADYKKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-L 331
              | ||| ||:|||:|   |      :: || ::: ||  |||::| | || :||  ::   :
Db        296 LNDYKAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIM 355

Qy        332 YSFTEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388
              | ||| :|  || :| | :|    :   |: |||:|:||| :||:||   :||  ::|||
Db        356 YGFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQN 415

Qy        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGEL 434
              :|     : | ||| |:|  || |:    ||::                :||:| | :|
Db        416 MRVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDL 475

Qy        435 FFVASENSYNDDNINTPKEIDDTVTSNN-----NYENDLDQVILNFNSESAPGLSDEKLN 489
              |:| :||:::::        |: |: |      |:   ||::|:::| :|    | :::
Db        476 TFIAEKNSFSEEPFQ-----DEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTT 530

Qy        490 LTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQP 549
                :  | |:| ||  | || |:::  :: || ||| |     : :|:|:| ||:
Db        531 PVTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINST 590

Qy        550 KIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLA 609
              |||::| |  |: ||:  |  ||: |:: :: ||| |::||:|:|||:|:| :|||||
Db        591 KIYSYFPS-VISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPA 649

Qy        610 LNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINN 669
              |||  :  :|||  |||  |  :|||: ||: :| |   :|     ||  | |:|| |:|
Db        650 LNIVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAE---SSTQKEKIIKTIDN 706

Qy        670 ALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLE 729
              |::| ||| |||  : : |:  :|||| ||  |||::|: ||:||| ||: :|  |:
Db        707 FLEKRYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGP 766

Qy        730 EKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLM-KLINEVKINKLREYDEN 788
              :| :: :  :|  ::|:| :| : || ||: |: ||| |:|: ::||| |  :| |:|
Db        767 DKEQIAD--EINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAK-KQLLEFDTQ 823

Qy        789 VKTYLLNYIIQHGSILGESQ-QELNSMVTDTLNNSIPFKLSSYTD---------DKILIS 838
              |  |: ||  :   :| :: ::| | :    :  |||  |   |          | ||
Db        824 SKNILMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL-- 881

Qy        839 YFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSE 897
                    | |::||:   ||   | ||::|::      |   |  |      : |:: ||
Db        882 ---------KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSE 932

Qy        898 VNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNV--NNEYTIINCMRDNN----SGWK 951

Db        933 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 992

Qy        952 VSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYI-NKWIFVTITNDRLGDSKLYINGN 1010
              |||  | :||||:|:||  ::: |     :  : |: |||:|:|||||||  : |||||
Db        993 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 1051

Qy       1011 LIDQKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069
              |:    |  || |   :||  |:  | :  :|: |  | || | |:  ||: ||::  :
Db       1052 LMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSI 1111

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLAN---------- 1119
                |:|||| | || |||| |   :      || : : ||   : | |
Db       1112 TFLRDFWGNPLRYDTEYYLIPVASSS------KD--VQLKNITDYMYLTNAPSYTNGKLN 1163

Qy       1120 ----RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLFPLYADTATTNKEK 1173
                   |||:|:|  |:|   ::  |: |:  | ::|:::                  | |
Db       1164 IYYRRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSY---------------NNNEH 1207

Qy       1174 TIKISSSGNRFNQVVVMNSVGNNC--------------------TMNFKNNNGNNIGL 1211
                :      || || :  :  || |                    :   ::   ::||
Db       1208 IVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGL 1267

Qy       1212 LGF---------KADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGW 1249
                :|          | ::|| ||: |::|     || | |:  : ||
Db       1268 VGTHNGQIGNDPNRDILIASNWYFNHLKD--KILGCDWYFVPTDEGW 1312


RESULT 6
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-20
```

```
  Query Match             32.7%;  Score 2141.5;  DB 4;  Length 1169;
  Best Local Similarity   39.2%;  Pred. No. 2.3e-131;
  Matches  466;  Conservative  250;  Mismatches  414;  Indels   59;  Gaps   26;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||::  |: ::|    |  :||:| |  |||||||  |  |:||:
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : | |  ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :  |  :|:  ||    |  :  ::     :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  ||| |  : | |||   ||:   |:         : |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :    :  ::   : :      ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |     |  :| ||: |: ||   :|: | ||::   |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|      |: |||:: || | ||:||:  :::  :|||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :   |:   |    |: ||   |||     ||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:  | :|  |||  ::::||: :  |   |  |         :  :| |:       |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
                :|  :| || :|  |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  ||| : :|:||||||::|||||||:|||  ||||::| |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              ||||| |||||| :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKV 751
              :||||   || ||:||  |  |:: ||: :|| |: :||: :   |  | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQEL 811
              : |::||: |:   |:|||| :  : : || ::|  :|  |||| ::   |  | :
Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
              | |   |  :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |    :|||| : :   |:: ::||  ||::: : :||:|||:||| | |: :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINK 989
              ::|||||||||: ||||||:|:  | ||||| |  |  : : | |     ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001
```

```
Qy        990 WIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | |||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS---- 1104
              ||: || ::  |:  |  :  :  |||||| |:|:||||: |    |::|  :|||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYI 1150
                | |  |  |  :    || | |  |:| :|| | ||::||| | :|:
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYL 1169
```

RESULT 7
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-22

```
  Query Match             23.3%;  Score 1526.5;  DB 4;  Length 858;
  Best Local Similarity   38.8%;  Pred. No. 2e-91;
  Matches  338;  Conservative  181;  Mismatches  308;  Indels   45;  Gaps   19;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||::  |: ::|    |  :||:| |  |||||||  |  |:||:
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : | |  ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :  |  :|:  ||    |  :  ::     :  |:|| |  | | |  : :| ::|:
```

```
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  |  ||| |  : | |||   ||:    |:          : |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :       ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |      |  :| ||: |: ||   :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|        |: |||:: || | ||:||:  :::  :|||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:   |    |: ||   |||    |||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  |||  ::::||: :   |   |         :  :| |:      |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
                :|  :| || :|  |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::||||||||:|||  ||||::|  |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              ||||| ||||||  :  |  ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKV 751
               :||||   || ||:||  |  |:: ||: :|| |: :||: :    |   | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQEL 811
              : |::||: |:   |:|||| :  : : || ::|  :|  |||| ::    |  | :
Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVTDTLNNSIPFKLSSYTDDKILISYFNKF 843
               | |   |   :|| || ||:| |||  |||:
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKY 856


RESULT 8
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             23.3%;  Score 1526.5;  DB 4;  Length 858;
  Best Local Similarity   38.8%;  Pred. No. 2e-91;
  Matches  338;  Conservative  181;  Mismatches  308;  Indels   45;  Gaps   19;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||:::  |: ::|    |   :||:| |  |||||||   |  |:||:
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : : | |   ||||:|| ::::|: ||: : |:|||| :   |  ||| :   ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   |  :|:   ||    |  :   ::      :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   ||:    |:          : |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :       ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |      |  :| ||: |: ||   :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
               ::|  :::| | :|        |: |||:: || | ||:||:  :::  :|||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:   |    |: ||   |||    |||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  |||  ::::||: :   |   |         :  :| |:      |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
                :|  :| || :|  |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::||||||||:|||  ||||::|  |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
```

```
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKV 751

Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQEL 811

Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVTDTLNNSIPFKLSSYTDDKILISYFNKF 843

Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKY 856


RESULT 9
US-09-255-829-18
; Sequence 18, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1013 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-18

  Query Match             20.8%;  Score 1359.5;  DB 4;  Length 1013;
  Best Local Similarity   34.3%;  Pred. No. 2e-80;
  Matches  352;  Conservative  198;  Mismatches  368;  Indels  107;  Gaps   36;

Qy          7 FNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62

Db          8 FNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122

Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180

Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231

Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285

Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341

Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398

Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 ITG-RGLVKKIIRFC-KNIVSVK----------------GIRKSICIEINNGELFFVASE 440

Db        416 LKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVNNWDLFFSPSE 475

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491

Db        476 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 530

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548

Db        531 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 586

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608

Db        587 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 646

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668

Db        647 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 703

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728

Db        704 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 763

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788

Db        764 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 821

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847

Db        822 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKSG 880

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPT-NKNQFGIYNDKLSEVNISQNDYI 906

Db        881 LNSPGAAHYAQHDEAVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKD-DPSQSANL 939

Qy        907 IYDNKYKNFSISFWVRIPNYDNKIVNVNNE-----YTIINCMRDNNSGWKVSLNHNEIIW 961

Db        940 LAEAKKLNDA-----QAPKVDNKF---NKEQQNAFYEILHLPNLNEE------QRNAFIQ 985

Qy        962 TLQDN 966

Db        986 SLKDD 990


RESULT 10
US-09-255-829-16
; Sequence 16, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
```

```
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-16

  Query Match             20.7%;  Score 1353;  DB 4;  Length 907;
  Best Local Similarity   35.8%;  Pred. No. 4.6e-80;
  Matches  333;  Conservative  182;  Mismatches  322;  Indels   94;  Gaps   31;

Qy          7 FNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | |||     |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::           :|  |||:|| |||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
                | |   ||    : |  :  ||  |||| |   | | | |:          :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 ITG-RGLVKKIIRFC-KNIVSVK---------------GIRKSICIEINNGELFFVASE 440
              :   || :     | : |:: |               |    :||::|| :|||  ||

Db        416 LKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVNNWDLFFSPSE 475

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||     ||: | |         : : |  | || :: |  :| | |:
Db        476 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 530

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |     :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        531 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 586

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
               ::||||||:::  |||  :||:|: |::|:: ||| | :: || |||||:|::||||
Db        587 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 646

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||   :: |:
Db        647 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 703

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        704 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 763

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        764 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 821

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRI 847
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: |
Db        822 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKSR 880

Qy        848 KSSSVLNMRYKNDKYVDTSGYDSNININGDV 878
              | |   :         ||| |:|: :  :
Db        881 PQSKVKRQIF--------SGYQSDIDTHNRI 903


RESULT 11
US-09-255-829-8
; Sequence 8, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
```

```
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-8

  Query Match             20.6%;  Score 1348;  DB 4;  Length 871;
  Best Local Similarity   36.6%;  Pred. No. 9.3e-80;
  Matches  331;  Conservative  178;  Mismatches  302;  Indels   94;  Gaps   32;

Qy          1 MPKIN-SFNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DF 55
              || :|  ||| ||||  | |||    |  :  |:| |   ||:||||:   | |:  |
Db          1 MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 HPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
              :||   |    ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
              :| | :  : | : : :   :||       |::|:|   |:  ||  |    : |
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
              : :|:||   : |||:::| | ::          :|  |||:|| |||||: | ||
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
              ||        | |   ||    : |  :  ||  |||| |    | | | |:
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
              :| ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
              || :    |:|  |:||:    |  ||: |::    | | :|:|:   ||  || |||  :
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
              |   | :   ||       :  | | :: |:||  |                 :||::|| :
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
              |||  ||::: :| :|  :||   ||: | |         : : |  | || :: |  :|
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              | |:  |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              : |||   ::|||||:::  |||  :||:|: |:::|:: ||| | :: || ||||||:|
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
              ::|||| ||||||   | :|  ||   || ||||| ||: || :  | : |::   ||
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIES 721
              ::  |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||   | | ||
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
              :|| || |||| :   ::|  : ::||: :: || ||::|| : |:|||| :    : :
Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              | ::| ::|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFK 845
```

```
              :: |
Db        867 TEYIK 871


RESULT 12
US-09-255-829-2
; Sequence 2, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-2

  Query Match             20.5%;  Score 1340.5;  DB 4;  Length 871;
  Best Local Similarity   36.5%;  Pred. No. 2.9e-79;
  Matches  328;  Conservative  177;  Mismatches  300;  Indels   93;  Gaps   31;

Qy          7 FNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | |||    |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                 ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||       |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::            :|  |||:|| |||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              | | ||    : |  :  ||  |||| |   | | | |:       :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296
```

```
Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :    |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS----------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:|| |                :||::|| :||| ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |        : : |  | || :: |  :| | |:
Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
               ::||||||:::  |||  :||:|: |::|:: ||| | :: || |||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | :|::   ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK 845
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK 871
```

RESULT 13
US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-26

Query Match 20.5%; Score 1340.5; DB 4; Length 871;
Best Local Similarity 36.5%; Pred. No. 2.9e-79;
Matches 328; Conservative 177; Mismatches 300; Indels 93; Gaps 31;

```
Qy          7 FNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | |||    |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||     |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::           :|  |||:|| |||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
                | |   ||    : | :  ||  |||| |   | | | |:        :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :    |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS----------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:|| |                :||::|| :||| ||
Db        416 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 468

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |        : : |  | || :: |  :| | |:
Db        469 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 523

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        524 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
               ::||||||:::  |||  :||:|: |::|:: ||| | :: || |||||:|::||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | :|::   ||   :: |:
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 696

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        697 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 756

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::| :
```

```
Db        757 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK 845
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: |
Db        815 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK 871


RESULT 14
US-09-255-829-10
; Sequence 10, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  875 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-10

  Query Match             20.5%;  Score 1340.5;  DB 4;  Length 875;
  Best Local Similarity   36.6%;  Pred. No. 2.9e-79;
  Matches  329;  Conservative  175;  Mismatches  303;  Indels   93;  Gaps   31;

Qy          7 FNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              ||| ||||   | |||    |  :  |:| |   ||:||||:   | |:  | :||   |
Db          8 FNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 66

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         67 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 125

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |::|:|   |:  ||  |    : |    : :|:||
Db        126 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 182

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::           :|  |||:|| |||||: | ||   ||
Db        183 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 239

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
                | |   ||    : |  :  ||  |||| |   | | | |:         :| |||
Db        240 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 296

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        297 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 356

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || |||  :|    |
Db        357 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 415

Qy        399 I---TG----------RGLVKKIIRFCKNIVSVKGIRKS-------ICIEINNGELFFVA 438
              :   ||          ||::       |     ||  ||       :||::|| :|||
Db        416 LKNFTGLFEFYKLLCVRGIITS-----KTKSLDKGYNKSADGALNDLCIKVNNWDLFFSP 470

Qy        439 SENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLN 489
              ||::: :| :|  :||     ||: | |         : : |  | || :: |  :| | |:
Db        471 SEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLS 525

Qy        490 LTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALL 546
                |      :|  :   ||    :::::::  :| || ||:   |:: : ||:|:: |||
Db        526 SDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALL 581

Qy        547 EQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYI 606
                ::|||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|::|||
Db        582 NPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYI 641

Qy        607 GLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKA 666
              | ||||||   | :|  ||   || ||||| ||: || :  | : |::    ||   ::
Db        642 GPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQT 698

Qy        667 INNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSY 726
              |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| |
Db        699 IDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQY 758

Qy        727 TLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYD 786
              | |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :| ::|
Db        759 TEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFD 816

Qy        787 ENVKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK 845
              ::|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: |
Db        817 ASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK 875


RESULT 15
US-09-255-829-6
; Sequence 6, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 873 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-6

Query Match 20.5%; Score 1338.5; DB 4; Length 873;
Best Local Similarity 36.5%; Pred. No. 3.9e-79;
Matches 328; Conservative 177; Mismatches 300; Indels 93; Gaps 31;

```
Qy          7 FNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFHPPTSLK 62
              |||:|||||   | |||    |  :  |:| |   ||:||||:   | |:  | :||   |
Db         10 FNYKDPVNGVDIAYIKIPKYGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPPEAK 68

Qy         63 NGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDN 122
                ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |  :| |
Db         69 QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG-STIDT 127

Qy        123 QFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPSNHGFGS 180
              :  : | : : :   :||      |::|:|   |:  ||  |    : |    : :|:||
Db        128 ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LTRNGYGS 184

Qy        181 IAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAKGITTKY 231
              : |||:::| | ::           :|  |||:|| |||||: | ||   ||
Db        185 TQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY---GIAINP 241

Qy        232 TITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
                | |   ||    : |  :  ||  |||| |    | | | |:         :| |||
Db        242 NRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIAST 298

Qy        286 LSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTEFDLAT 341
              |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|| :
Db        299 LNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVK 358

Qy        342 KFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNPRIITP 398
              |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :|    |
Db        359 FFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTK 417

Qy        399 I---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGELFFVASE 440
              :   ||      :  | | :: |:||  |              :||::|| :|||  ||
Db        418 LKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSE 470

Qy        441 NSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSDEKLNLT 491
              ::: :| :|  :||    ||: | |        : : |  | || :: |  :| | |:
Db        471 DNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSD 525

Qy        492 IQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 548
              |     :| :   ||    ::::::  :| || ||:   |:: : ||:|:: |||
Db        526 IIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 581

Qy        549 PKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 608
               ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|::||||
Db        582 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 641

Qy        609 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 668
              ||||||   | :|  ||   || ||||| ||: || :  | : |::   ||   :: |:
Db        642 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTID 698

Qy        669 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTL 728
              ||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :|| ||
Db        699 NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE 758

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 788
              |||| :   ::|  : ::||: :: || ||::|| : |:|||| :    : :| ::| :
Db        759 EEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 816

Qy        789 VKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK 845
              :|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| | :: |
Db        817 LKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK 873
```

Search completed: September 10, 2004, 03:47:04
Job time : 30.574 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 242.348 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 6541
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6541	100.0	1252	12	US-10-272-898-52	Sequence 52, Appl
2	6541	100.0	1252	15	US-10-354-774-52	Sequence 52, Appl
3	6541	100.0	1252	15	US-10-271-012-52	Sequence 52, Appl
4	6541	100.0	1252	15	US-10-452-024-128	Sequence 128, App
5	6541	100.0	1252	16	US-10-729-122-52	Sequence 52, Appl
6	6541	100.0	1252	16	US-10-729-039-52	Sequence 52, Appl
7	6492.5	99.3	1251	15	US-10-452-024-129	Sequence 129, App
8	6361	97.2	1252	15	US-10-452-024-125	Sequence 125, App
9	6361	97.2	1252	15	US-10-205-516-10	Sequence 10, Appl
10	6361	97.2	1255	15	US-10-452-024-126	Sequence 126, App
11	6349	97.1	1262	15	US-10-205-516-24	Sequence 24, Appl
12	6344.5	97.0	1251	15	US-10-452-024-5	Sequence 5, Appli
13	6344.5	97.0	1251	15	US-10-452-024-124	Sequence 124, App
14	6342	97.0	1250	12	US-10-272-898-50	Sequence 50, Appl
15	6342	97.0	1250	15	US-10-354-774-50	Sequence 50, Appl
16	6342	97.0	1250	15	US-10-271-012-50	Sequence 50, Appl
17	6342	97.0	1250	16	US-10-729-122-50	Sequence 50, Appl
18	6342	97.0	1250	16	US-10-729-039-50	Sequence 50, Appl
19	6335.5	96.9	1251	15	US-10-452-024-127	Sequence 127, App
20	4296.5	65.7	1278	15	US-10-452-024-152	Sequence 152, App
21	4296.5	65.7	1278	15	US-10-205-516-12	Sequence 12, Appl
22	4284.5	65.5	1288	15	US-10-205-516-26	Sequence 26, Appl
23	4204.5	64.3	1268	15	US-10-452-024-156	Sequence 156, App
24	4153	63.5	804	13	US-10-011-588-37	Sequence 37, Appl
25	4077	62.3	1280	15	US-10-452-024-162	Sequence 162, App
26	4056.5	62.0	1274	12	US-10-272-898-71	Sequence 71, Appl
27	4056.5	62.0	1274	15	US-10-354-774-71	Sequence 71, Appl
28	4056.5	62.0	1274	15	US-10-271-012-71	Sequence 71, Appl
29	4056.5	62.0	1274	15	US-10-452-024-6	Sequence 6, Appli
30	4056.5	62.0	1274	16	US-10-729-122-71	Sequence 71, Appl
31	4056.5	62.0	1274	16	US-10-729-039-71	Sequence 71, Appl
32	2728.5	41.7	858	13	US-10-011-588-41	Sequence 41, Appl
33	2390	36.5	449	10	US-09-910-186A-14	Sequence 14, Appl
34	2345	35.9	1296	15	US-10-452-024-160	Sequence 160, App
35	2345	35.9	1296	15	US-10-205-516-2	Sequence 2, Appli
36	2344	35.8	1296	15	US-10-452-024-159	Sequence 159, App
37	2343	35.8	1306	15	US-10-205-516-16	Sequence 16, Appl
38	2339	35.8	1295	9	US-09-726-949A-1	Sequence 1, Appli
39	2339	35.8	1295	15	US-10-452-024-157	Sequence 157, App
40	2336.5	35.7	1296	12	US-10-272-898-28	Sequence 28, Appl
41	2336.5	35.7	1296	15	US-10-354-774-28	Sequence 28, Appl
42	2336.5	35.7	1296	15	US-10-271-012-28	Sequence 28, Appl
43	2336.5	35.7	1296	15	US-10-452-024-158	Sequence 158, App
44	2336.5	35.7	1296	16	US-10-729-122-28	Sequence 28, Appl
45	2336.5	35.7	1296	16	US-10-729-039-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-52
; Sequence 52, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:

```
;           LENGTH: 1252 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-272-898-52

  Query Match            100.0%;  Score 6541;  DB 12;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 2
US-10-354-774-52
; Sequence 52, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1252 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-354-774-52

  Query Match            100.0%;  Score 6541;  DB 15;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
```

```
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252

Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
```

RESULT 3
US-10-271-012-52
; Sequence 52, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-271-012-52

Query Match 100.0%; Score 6541; DB 15; Length 1252;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
```

```
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
```

RESULT 4

```
US-10-452-024-128
; Sequence 128, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 128
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-128

  Query Match           100.0%;  Score 6541;  DB 15;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
```

```
Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
```

RESULT 5
US-10-729-122-52
; Sequence 52, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-729-122-52

Query Match 100.0%; Score 6541; DB 16; Length 1252;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
```

```
Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 6
US-10-729-039-52
; Sequence 52, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 52:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1252 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-729-039-52

  Query Match           100.0%;  Score 6541;  DB 16;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
```

```
Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252


RESULT 7
US-10-452-024-129
; Sequence 129, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 129
;   LENGTH: 1251
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-129

  Query Match             99.3%;  Score 6492.5;  DB 15;  Length 1251;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1244;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHRFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              || :|| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-NCTMN 1199

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1251


RESULT 8
US-10-452-024-125
; Sequence 125, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 125
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-125
```

```
  Query Match             97.2%;  Score 6361;  DB 15;  Length 1252;
  Best Local Similarity   97.4%;  Pred. No. 0;
  Matches 1219;  Conservative   15;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              |||||||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              |||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1252


RESULT 9
US-10-205-516-10
; Sequence 10, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 1252
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-10

  Query Match             97.2%;  Score 6361;  DB 15;  Length 1252;
  Best Local Similarity   97.4%;  Pred. No. 0;
  Matches 1219;  Conservative   15;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||| ||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1252


RESULT 10
US-10-452-024-126
; Sequence 126, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 126
;   LENGTH: 1255
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-126

  Query Match             97.2%;  Score 6361;  DB 15;  Length 1255;
  Best Local Similarity   97.4%;  Pred. No. 0;
  Matches 1219;  Conservative   15;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||:|||||||||||||||||||| |||| ||||
Db          4 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 63

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||:|||:||||||||||||:|||| ||||||||||||||||||||
Db         64 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 123

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        124 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 183

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        184 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 243

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 303

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 363

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 423

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 483

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 543

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        544 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 603

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 663

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        664 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 723

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db        724 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 783

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||| ||||||||||||||||||||||||||
Db        784 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 843

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 903

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 963

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        964 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1023

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
```

```
                ||||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||
Db       1024 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1083

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
                ||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1084 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1143

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1144 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1203

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
                |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1204 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1255


RESULT 11
US-10-205-516-24
; Sequence 24, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 1262
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-24

  Query Match             97.1%;  Score 6349;  DB 15;  Length 1262;
  Best Local Similarity   97.1%;  Pred. No. 0;
  Matches 1219;  Conservative   15;  Mismatches   18;  Indels    4;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 I----RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFN 476
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRIEGRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFN 480

Qy        477 SESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVN 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVN 540

Qy        537 LTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKI 596
              ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        541 LTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKI 600

Qy        597 ADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGS 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGS 660

Qy        657 SDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIK 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        661 SDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALK 720

Qy        717 TIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINE 776
              |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        721 AIIESKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINE 780

Qy        777 VKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKIL 836
              |||||||||||||||||:|||:|||||||||||||| |||||||||||||||||||||||
Db        781 VKINKLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKIL 840

Qy        837 ISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS 900

Qy        897 EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNH 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNH 960

Qy        957 NEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKS 1016
              ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||:||
Db        961 NEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKS 1020

Qy       1017 ILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFW 1076
              ||||||||||||||||||||||||||||||||||||||||||||||:|||| |||||||
Db       1021 ILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFW 1080

Qy       1077 GNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS 1136
              |||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 GNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS 1140

Qy       1137 TNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNN 1196
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1141 TNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNN 1200

Qy       1197 CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 CTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1256


RESULT 12
US-10-452-024-5
; Sequence 5, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
;   LENGTH: 1251
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-5
```

```
  Query Match           97.0%;  Score 6344.5;  DB 15;  Length 1251;
  Best Local Similarity  97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||||:||||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||| ||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-NCTMN 1199

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1200 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1251


RESULT 13
US-10-452-024-124
; Sequence 124, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 124
;   LENGTH: 1251
;   TYPE: PRT
;   ORGANISM: Clostridium butyricum
US-10-452-024-124

  Query Match           97.0%;  Score 6344.5;  DB 15;  Length 1251;
  Best Local Similarity  97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||||:||||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
```

```
Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-NCTMN 1199

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
Db       1200 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1251


RESULT 14
US-10-272-898-50
; Sequence 50, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1250 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-272-898-50

  Query Match             97.0%;  Score 6342;  DB 12;  Length 1250;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 1218;  Conservative   15;  Mismatches   17;  Indels    2;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
```

```
Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||| ||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||| |||||||||||||||||||||||||||||||| ||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
```

RESULT 15
US-10-354-774-50
; Sequence 50, Application US/10354774
; Publication No. US20030215468A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/354,774
; FILING DATE: 30-Jan-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1250 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-354-774-50

Query Match 97.0%; Score 6342; DB 15; Length 1250;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1218; Conservative 15; Mismatches 17; Indels 2; Gaps 1;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:||||||||||||:||||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||||:|||:||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||| ||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840
```

```
Qy       841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy       901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy       961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
            ||||||:||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db       961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy      1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
            ||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||||
Db      1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy      1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
            ||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||||
Db      1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy      1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
            |||||||||||||||||| ||||||||||||||||||||||||||||||  ||||||||||||
Db      1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG--CTMN 1198

Qy      1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
            ||||||||||||||||||||||||||||:||||| ||||||||||||||||||
Db      1199 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
```

Search completed: September 10, 2004, 04:58:10
Job time : 247.348 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 22.1185 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 6541
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6541	100.0	1252	2	S21178	botulinum neurotox
2	6344.5	97.0	1251	2	JH0256	botulinum neurotox
3	4204.5	64.3	1268	2	S33411	botulinum neurotox
4	4056.5	62.0	1274	2	I40813	neurotoxin type F
5	2345	35.9	1296	1	BTCLAB	bontoxilysin (EC 3
6	2290	35.0	1296	2	I40645	botulinum neurotox
7	2257	34.5	1297	2	S39791	neurotoxin - Clost
8	2208	33.8	1291	2	I40631	non-proteolytic bo
9	2204	33.7	1291	1	A48940	bontoxilysin (EC 3
10	2151.5	32.9	1315	1	BTCLTN	tentoxilysin (EC 3
11	1895	29.0	367	2	S48106	neurotoxin type E
12	1883	28.8	1276	2	S11455	botulinum neurotox
13	1863	28.5	1285	2	S70582	botulinum neurotox
14	1822	27.9	1291	2	S46431	botulinum neurotox
15	1822	27.9	1291	2	A49777	botulinum neurotox
16	1425.5	21.8	366	2	S48110	neurotoxin type F
17	1314	20.1	369	2	S48109	neurotoxin type F
18	762.5	11.7	1196	2	S46430	botulinum neurotox
19	761.5	11.6	1196	2	JQ1467	toxin, nontoxic co
20	754	11.5	1193	2	JC4901	nontoxic-nonhemagg
21	747.5	11.4	1162	2	A47708	progenitor toxin n
22	734.5	11.2	1162	2	I40817	botulinum toxin no
23	685.5	10.5	1193	2	S68218	botulinum neurotox
24	667.5	10.2	552	2	A49928	neurotoxin C1, BoN
25	667.5	10.2	1165	2	I40644	botulinum neurotox
26	443	6.8	122	2	A53878	type E neurotoxin
27	359.5	5.5	1436	2	D71618	hypothetical prote
28	346	5.3	2401	2	T28676	rhoptry protein -
29	340	5.2	1711	2	T18429	hypothetical prote
30	333	5.1	2269	2	T28677	rhoptry protein -
31	330.5	5.1	3724	2	T18427	hypothetical prote
32	328.5	5.0	2485	1	H71621	serine/threonine-s
33	324.5	5.0	1127	2	T28317	ORF MSV156 hypothe
34	322.5	4.9	4688	2	F82885	hypothetical prote
35	321	4.9	1817	2	D71606	hypothetical prote
36	319	4.9	2380	2	E71604	hypothetical prote
37	316.5	4.8	1844	2	D71612	hypothetical prote
38	316.5	4.8	4981	2	T18489	hypothetical prote
39	315.5	4.8	2539	2	B71619	hypothetical prote
40	315	4.8	1650	2	T18444	hypothetical prote
41	314.5	4.8	3973	2	B71612	hypothetical prote
42	314.5	4.8	4550	2	T18440	hypothetical prote
43	311	4.8	3394	2	T18501	hypothetical prote
44	310.5	4.7	3844	2	T18402	asparagine/asparta
45	309.5	4.7	1306	2	T28313	ORF MSV152 probabl

ALIGNMENTS

RESULT 1
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin

```
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match           100.0%;  Score 6541;  DB 2;  Length 1252;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-259;
  Matches 1252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252

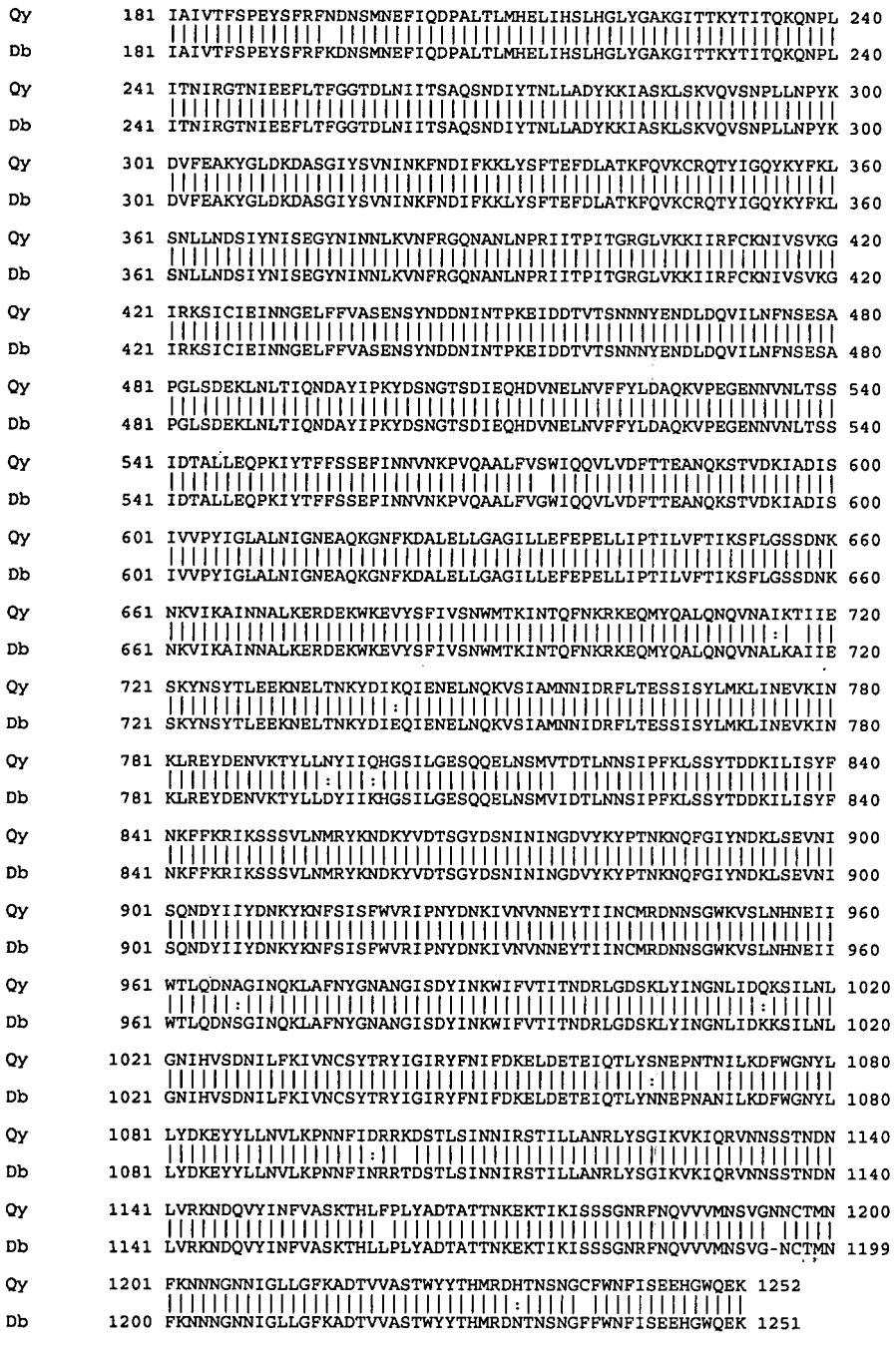

RESULT 2
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.;
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxir
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release a
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match           97.0%;  Score 6344.5;  DB 2;  Length 1251;
  Best Local Similarity  97.3%;  Pred. No. 1.8e-251;
  Matches 1218;  Conservative   15;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

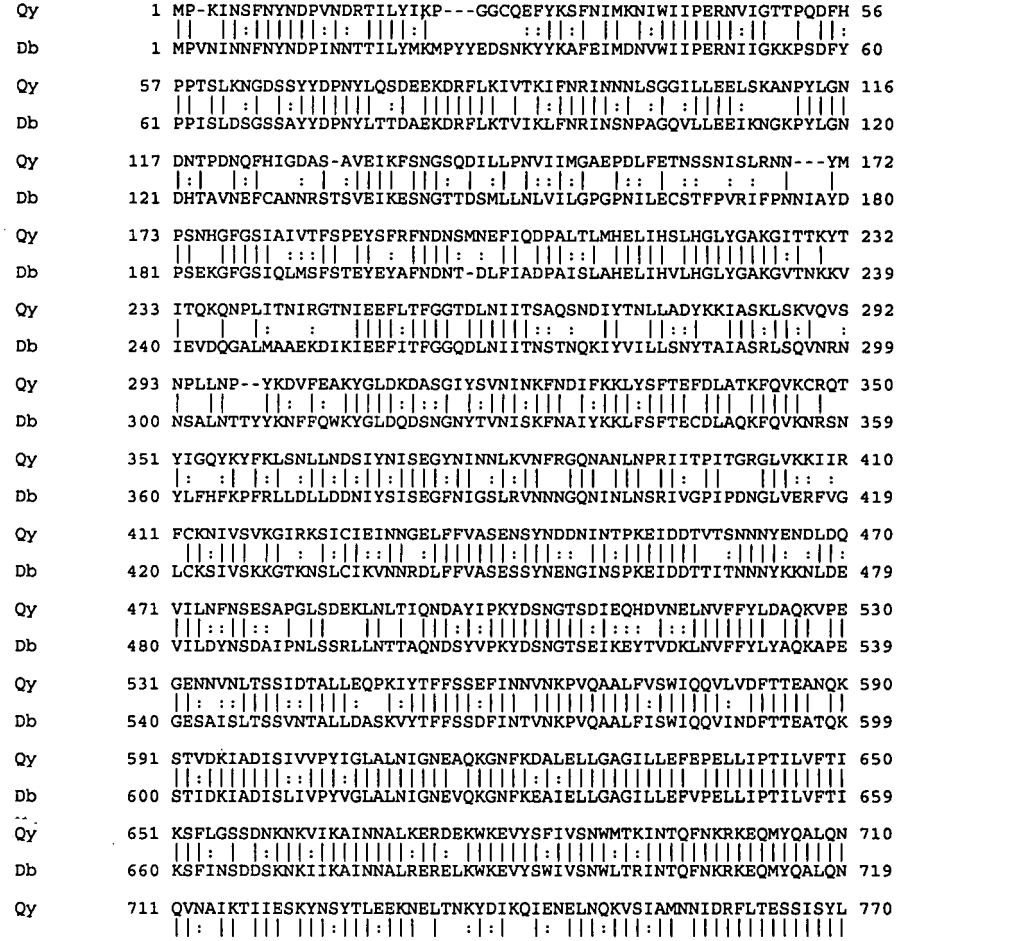
Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
```



RESULT 3
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 64.3%; Score 4204.5; DB 2; Length 1268;
Best Local Similarity 63.1%; Pred. No. 3.5e-164;
Matches 803; Conservative 219; Mismatches 225; Indels 25; Gaps 14;



```
Db        720 QVDGIKKIIEYKYNNYTLDEKNRLRAEYNIYSIKEELNKKVSLAMQNIDRFLTESSISYL 779

Qy        771 MKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQ-QELNSMVTDTLNNSIPFKLSS 829

Db        780 MKLINEAKINKLSEYDKRVNQYLLNYILENSSTLGTSSVPELNNLVSNTLNNSIPFELSE 839

Qy        830 YTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFG 889

Db        840 YTNDKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFG 899

Qy        890 IYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSG 949

Db        900 IYSSRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN-LKNLNNEYTIINCMRNNNSG 958

Qy        950 WKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYING 1009

Db        959 WKISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYING 1018

Qy       1010 NLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Db       1019 NLTDQKSILNLGNIHVDDNILFKIVGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIR----STILLAN-RLYSG 1124

Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPNMSVTKNSD-ILNINRQRGIYSKTNIFSNARLYTG 1137

Qy       1125 IKVKIQRVNNSSTN--DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIK---ISS 1179

Db       1138 VEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSE-YQLYADVSTSAVEKTIKLRRISN 1196

Qy       1180 SGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCF 1239

Db       1197 SNYNSNQMIIMDSIGDNCTMNFKTNNGNDIGLLGFHLNNLVASSWYYKNIRNNTRNNGCF 1256

Qy       1240 WNFISEEHGWQE 1251

Db       1257 WSFISKEHGWQE 1268


RESULT 4
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            62.0%;  Score 4056.5;  DB 2;  Length 1274;
  Best Local Similarity  61.0%;  Pred. No. 3.9e-158;
  Matches  781;  Conservative  221;  Mismatches  242;  Indels   37;  Gaps   16;

Qy          1 MP-KINSFNYNDPVNDRTILYIK---PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTNPSDFD 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116

Db         61 PPASLKNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYLGN 120

Qy        117 DNTPDNQFH-IGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRN------ 169

Db        121 DHTPIDEFSPVTRTTSVNIKLSTNVESSMLLNLLVLGAGPDIFE--SCCYPVRKLIDPDV 178

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFND------NSMNEFIQDPALTLMHELIHSLHGLYG 223

Db        179 VYDPSNYGFGSINIVTFSPEYEYTFNDISGGHNSSTESFIADPAISLAHELIHALHGLYG 238

Qy        224 AKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIA 283

Db        239 ARGVTYEETIEVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIA 298

Qy        284 SKLSKVQVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLAT 341

Db        299 TRLSEVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTESDLAN 358

Qy        342 KFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITG 401

Db        359 KFKVKCRNTYFIKYEFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQSIKLNPKIIDSIPD 418

Qy        402 RGLVKKIIRFCKNIVSVKGIR--KSICIEINNGELFFVASENSYNDDNINTPKEIDDTVT 459

Db        419 KGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNSELFFVASESSYNENDINTPKEIDDTTN 478

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519

Db        479 LNNNYRNNLDEVILDYNSQTIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYDVVDFNV 538

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 579

Db        539 FFYLHAQKVPEGETNISLTSSIDTALLEESK-DIFFSSEFIDTINKPVNAALFIDWISKV 597

Qy        580 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPE 639

Db        598 IRDFTTEATQKSTVDKIADISLIVPYVGLALNIIIEAEKGNFEEAFELLGVGILLEFVPE 657

Qy        640 LLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNK 699

Db        658 LTIPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRINTQFNK 717

Qy        700 RKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNID 759

Db        718 RKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESEYNINNIEEELNKKVSLAMKNIE 777

Qy        760 RFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTL 819

Db        778 RFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLLNYILDHRSILGEQTNELSDLVTSTL 837

Qy        820 NNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVY 879

Db        838 NSSIPFELSSYTNDKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVY 897

Qy        880 KYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTI 939

Db        898 IYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTI 956

Qy        940 INCMRDNNSGWKVSL---NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTIT 996

Db        957 INCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTIT 1016

Qy        997 NDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDE 1056

Db       1017 NNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDK 1076

Qy       1057 TEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIR---- 1112

Db       1077 TEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYI-TLNSGILNINQQRGVTE 1135
```

```
Qy       1113 STILLANRLYSGIKVKIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTN 1170
Db       1136 GSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSE 1193

Qy       1171 KEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMR 1230
Db       1194 KEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIR 1253

Qy       1231 DHTNSNGCFWNFISEEHGWQE 1251
Db       1254 RNTSSNGCFWSSISKENGWKE 1274
```

RESULT 5
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 35.9%; Score 2345; DB 1; Length 1296;
Best Local Similarity 40.0%; Pred. No. 2.6e-88;
Matches 536; Conservative 248; Mismatches 421; Indels 136; Gaps 43;

```
Qy          1 MPKIN-SFNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DF 55
Db          1 MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 HPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173
Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224
Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278
Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334
Db        290 FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 391
Db        350 TEDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI 408

Qy        392 NPRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS---------------ICIEINNGE 433
Db        409 NNMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD 461

Qy        434 LFFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLS 484
Db        462 LFFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENIS 516

Qy        485 DEKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
Db        517 IENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 572

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
Db        573 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 632

Qy        602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
```

```
Db        633 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKV 689

Qy        662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIES 721

Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        722 KYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781

Db        750 QYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKR 807

Qy        782 LREYDENVKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840

Db        808 LEDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTF 866

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Db        867 TEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEV 926

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Db        927 ILKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEII 984

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020

Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1021 GNIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNY 1079

Db       1045 GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDY 1104

Qy       1080 LLYDKEYYLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYS 1123

Db       1105 LQYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSS 1180

Db       1157 GTKFIIKKY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy       1181 GNRFNQVVVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRD 1231

Db       1215 GN-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIER 1273

Qy       1232 HTNSNGCFWNFISEEHGWQEK 1252

Db       1274 SSRTLGCSWEFIPVDDGWGER 1294


RESULT 6
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             35.0%;  Score 2290;  DB 2;  Length 1296;
  Best Local Similarity   38.7%;  Pred. No. 4.6e-86;
  Matches  523;  Conservative  240;  Mismatches  429;  Indels  160;  Gaps   38;

Qy          1 MPKIN-SFNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DF 55

Db          1 MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDL 59

Qy         56 HPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115

Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMP 173

Db        120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHDVLN---L 175

Qy        174 SNHGFGSIAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGA 224

Db        176 TRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAEHRLY-- 233

Qy        225 KGITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLAD 278

Db        234 -GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK 289

Qy        279 YKKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-F 334

Db        290 FKDVASTLNKAKSIIGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIY 349

Qy        335 TEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLN 392

Db        350 TEDNFVNFFKVINRKTYLNFDKAVFRINIVPDENYTIKDGFNLKGANLSTNFNGQNTEIN 409

Qy        393 PRIITPI---TGRGLVKKIIRFCKNIVSVKGI--------------RKSICIEINNGEL 434

Db        410 SRNFTRLKNFTG------LFEFYK-LLCVRGIIPFKTKSLDEGYNKALNDLCIKVNNWDL 462

Qy        435 FFVASENSYNDDNINTPKEID--DTVTSNNNYE--------NDLDQVILNFNSESAPGLS 484

Db        463 FFSPSEDNFTND-------LDKVEEITADTNIEAAEENISLDLIQQYYLTFDFDNEPE-- 513

Qy        485 DEKLNLTIQNDAYIPKYDSNGTSDI----------------EQHDVNELNVFFYLDAQKV 528

Db        514 ----NISIE----------NLSSDIIGQLEPMPNIERFPNGKKYELDKYTMFHYLRAQEF 559

Qy        529 PEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEAN 588

Db        560 EHGDSRIILTNSAEEALLKPNVAYTFFSSKYVKKINKAVEAFMFLNWAEELVYDFTDETN 619

Qy        589 QKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVF 648

Db        620 EVTTMDKIADITIIVPYIGPALNIGNMLSKGEFVEAIIFTGVVAMLEFIPEYALPVFGTF 679

Qy        649 TIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQAL 708

Db        680 AIVSYIA---NKVLTVQTINNALSKRNEKWDEVYKYTVTNWLAKVNTQIDLIREKMKKAL 736

Qy        709 QNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSIS 768

Db        737 ENQAEATKAIINYQYNQYTEEEKNNI--NFNIDDLSSKLNESINSAMININKFLDQCSVS 794

Qy        769 YLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLS 828

Db        795 YLMNSMIPYAVKRLKDFDASVRDVLLKYIYDNRGTLVLQVDRLKDEVNNTLSADIPFQLS 854

Qy        829 SYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF 888

Db        855 KYVDNKKLLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQI 914

Qy        889 GIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNS 948

Db        915 KLINLESSTIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKI-NLNNEYTIINCI-ENNS 972

Qy        949 GWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYIN 1008

Db        973 GWKVSLNYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYIN 1032

Qy       1009 GNLIDQKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYSNEP 1067

Db       1033 GRLIDQKPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQS 1092

Qy       1068 NTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNI---------------- 1111
```

```
Db       1093 NSGILKDFWGNYLQYDKPYYMLNLFDPNKYVD--------VNNIGIRGYMYLKGPRGSVV 1144

Qy       1112 RSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNK 1171

Db       1145 TTNIYLNSTLYEGTKFIIKKY-ASGNEDNIVRNNDRVYINVVV-KNKEYRLATNASQAGV 1202

Qy       1172 EK---TIKISSSGNRFNQVVVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTV 1219

Db       1203 EKILSALEIPDVGN-LSQVVVMKSKDDQGIRNKCKMNLQDNNGNDIGFIGFHLYDNIAKL 1261

Qy       1220 VASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1251

Db       1262 VASNWYNRQVGKASRTFGCSWEFIPVDDGWGE 1293


RESULT 7
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            34.5%;  Score 2257;  DB 2;  Length 1297;
  Best Local Similarity  38.5%;  Pred. No. 1e-84;
  Matches  513;  Conservative  253;  Mismatches  449;  Indels  118;  Gaps   35;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115

Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        116 NDNTPDNQF--HIGDASAVEIKFSNGSQDI---LLPNVIIMGAEPDLFETNSSNISLRNN 170

Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNG 179

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221

Db        180 HSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239

Qy        222 YGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275

Db        240 YGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKA 292

Qy        276 LADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL- 331

Db        293 LQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALM 352

Qy        332 YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388

Db        353 FGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412

Qy        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNI 448

Db        413 KAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLA 472

Qy        449 NTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYD 502

Db        473 KA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYI 527

Qy        503 SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINN 562

Db        528 KQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEK 585

Qy        563 VNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFK 622

Db        586 ANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFK 645

Qy        623 DALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVY 682

Db        646 NAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMY 702

Qy        683 SFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ 742

Db        703 GLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFND 760

Qy        743 IENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGS 802

Db        761 IDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNEL 820

Qy        803 ILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKY 862

Db        821 YLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRL 880

Qy        863 VDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVR 922

Db        881 IDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVR 940

Qy        923 IPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNA 980

Db        941 TPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIK 999

Qy        981 NGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCS-YT 1039

Db       1000 DNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTT 1059

Qy       1040 RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFI- 1098

Db       1060 KFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK 1119

Qy       1099 ---DRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN-- 1151

Db       1120 YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 1179

Qy       1152 -----------FVASK---THLF-------PLYADTATTNK--EKTIKISSSGNRFNQVV 1188

Db       1180 NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKT--------TYNCQI 1231

Qy       1189 VMNSVGNNCTMNFKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDHTNS--NGC 1238

Db       1232 L-------CEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGC 1284

Qy       1239 FWNFISEEHGWQE 1251

Db       1285 NWQFIPVDEGWTE 1297


RESULT 8
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
```

```
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match             33.8%;  Score 2208;  DB 2;  Length 1291;
  Best Local Similarity   37.6%;  Pred. No. 1e-82;
  Matches  496;  Conservative  272;  Mismatches  454;  Indels   98;  Gaps   33;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||||:::  |: ::|     |   :||:| |   |||||||    |  |:||:
Db          1 MPVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : : | |   ||||:|| ::::|: | : : |:|||| :   |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFFQTLIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   |  :|:   ||    | :   ::     :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   ||:   |:          :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITN-----IRGTN---IEEFLTFGGTDLNIITSAQSNDIYT 273
              || |             |:: |     :: |:    ||  |||| | :||: :     ||
Db        240 YGIK---------VDDLPIVPNEKKFFMQSTDTIQAEELYTFGGQDPSIISPSTDKSIYD 290

Qy        274 NLLADYKKIASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFK 329
              :| ::: |  :|:|| |      |  :| ||: |: ||   :|: | ||:::  || ::|
Db        291 KVLQNFRGIVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFNKLYK 350

Qy        330 KL-YSFTEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFR 385
              |   ||| ::|  :::| | :|        |: |||:: || | ||:||:  |:   :|
Db        351 SLMLGFTEININAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMGKEYR 410

Qy        386 GQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYND 445
              |||  :| :    |:   |    |: ||: | | |    |||:::|  |||:| :||::|
Db        411 GQNKAINKQAYEEISKEHLAVYKIQMCKS-VKVPG----ICIDVDNENLFFIADKNSFSD 465

Qy        446 DNINTPKEIDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDS 503
              |   :  |  :  | ||   ||  :::||: :  |   |  |       :  :| |:
Db        466 D--LSKNERVEYNTQNNYIGNDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEK 523

Qy        504 NGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNV 563
                |::   :|  :| || :|  |    :::|||| | |||   |:|:||| ::|
Db        524 QPA--IKKVFTDENTIFQYLYSQTFPLNIRDISLTSSFDDALLVSSKVYSFFSMDYIKTA 581

Qy        564 NKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKD 623
              || |:| ||  |::|:: ||  |||: ||:||||||::|||||||||:|:|  ||||:
```

```
Db        582 NKVVEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGDETAKGNFES 641

Qy        624 ALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYS 683
              | |: |: ||||| |||||| : || ::|::   |||||:|| |:||| :| ||| ::|
Db        642 AFEIAGSSILLEFIPELLIPVVGVFLLESYI---DNKNKIIKTIDNALTKRVEKWIDMYG 698

Qy        684 FIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQI 743
              ||: |:: :|||   || ||:||  |  |:: ||: ||| |: |||: :   ::   |
Db        699 LIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNIYSEEEKSNININFN--DI 756

Qy        744 ENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSI 803
              ::||  :: ||:||: |: | |:||||| :  : : || ::|  :|  |||||  ::
Db        757 NSKLNDGINQAMDNINDFINECSVSYLMKKMIPLAVKKLLDFDNTLKKNLLNYIDENKLY 816

Qy        804 LGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYV 863
              |  | ::  | |   |   ||| ||:|:: :||| |||:   | :: :|||:||:::  :
Db        817 LIGSVEDEKSKVDKYLKTIIPFDLSTYSNIEILIKIFNKYNSEILNNIILNLRYRDNNLI 876

Qy        864 DTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRI 923
              | ||| : : :   |      :|||| : :   |:: ::||  ||::: : :||:|||:||
Db        877 DLSGYGAKVEVYDGVKL--NDKNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRI 934

Qy        924 PNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNAN 981
              | | |  :   ::|||||||||: ||||||:|:  | ||||| |  |  : : | |
Db        935 PKYRNDDIQNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIRE 993

Qy        982 GISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTR 1040
              ||:|||:| ||||||: | ::|:|||| |      | ::| : |:  | |||:  :   |:
Db        994 DISEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQ 1052

Qy       1041 YIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDR 1100
              :| ::||:||: :|::: |: :|  :  :  ||||||| |:|:||||: |     |::|
Db       1053 FIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKL 1112

Qy       1101 RKDSTL------SINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFV 1153
                |||::      |  |  |  :    || | |  |:| :|| | ||::||| | :::: |
Db       1113 VKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLV 1172

Qy       1154 ASKTHLFPLYADTATTNKEKTIKIS--SSGNRFNQVVVMNSVGN----NCTMNFKNN--N 1205
                  : :||      :|: : :|  |  | | : : :          :| : || :  :
Db       1173 LHHEE-WRVYAYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEES 1231

Qy       1206 GNNIGLLGFK------------ADTVVASTWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251
                ::|||:|             |    | ||   :  : : :: || | || :: || |
Db       1232 TDDIGLIGIHRFYESGVLRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291
```

```
RESULT 9
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publicat
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
```

```
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match            33.7%;  Score 2204;  DB 1;  Length 1291;
  Best Local Similarity  37.5%;  Pred. No. 1.5e-82;
  Matches  492;  Conservative  271;  Mismatches  467;  Indels   82;  Gaps   32;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||:::  |: ::|    |  :||:| |   |||||||   |  |:||:
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : : | |   ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||

Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   | :|:   ||    |  :   ::     :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   ||:   |:          :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     : ::  : :      ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |     |  :| ||: |: ||   :|: | ||::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
              ::|  :::| | :|        |: |||:: || | ||:||:  :::  :|||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              :    |:   |    |: ||   |||     |||:::| :|||:| :||::|| ::  :
Db        419 QAYEEISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNER 472

Qy        454 IDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQ 511
              |:   | :|  |||  ::::||: :  |   |  |          :  :| |:      |::
Db        473 IEYN-TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKK 529

Qy        512 HDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL 571
              :|  :| || :|   |    :::|||| | |||   |:|:||| ::|   || |:| |
Db        530 IFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGL 589

Qy        572 FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAG 631
              |  |::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::| |: ||
Db        590 FAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGAS 649

Qy        632 ILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMT 691
              ||||| |||||| :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |::
Db        650 ILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLS 706

Qy        692 KINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKV 751
              :||||   || ||:||  |  |:: ||: :|| |: :||: :   |   | ::||: :
Db        707 TVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGI 764

Qy        752 SIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQEL 811
              : |::||: |:   |:|||| :  : : || ::|  :|  |||| ::   |  | :
Db        765 NQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYE 824

Qy        812 NSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSN 871
              | |   |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| :
Db        825 KSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAK 884

Qy        872 ININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIV 931
              : :   |     :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :
Db        885 VEVYDGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGI 942

Qy        932 N--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINK 989
              ::|||||||||: ||||||:|:  | |||||  |  |  : : | |      ||:|||:
Db        943 QNYIHNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINR 1001

Qy        990 WIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFN 1048
              | ||||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:
Db       1002 WFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFS 1060

Qy       1049 IFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS---- 1104
              ||: || :: |:  |  :  :  |||||||  |:|:|||| :    |::|  :|||
Db       1061 IFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE 1120

Qy       1105 --TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFP 1161
                | |  |  |  :   || | |  |:| :|| | ||::||| | :|::|  :   :
Db       1121 ILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWR 1179
```

```
Qy       1162 LYADTATTNKEKTI---KISSSGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLG 1213
Db       1180 VYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIG 1239

Qy       1214 FK------------ADTVVASTWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251
Db       1240 IHRFYESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 10
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio

C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized,
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of syr
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptc
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 32.9%; Score 2151.5; DB 1; Length 1315;
Best Local Similarity 35.8%; Pred. No. 2.1e-80;
Matches 490; Conservative 260; Mismatches 444; Indels 173; Gaps 37;

```
Qy          1 MP-KINSFNYNDPVNDRTILYIKPGGCQE---FYKSFNIMKNIWIIPERNVIGTTPQDFH 56
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        117 DNTPDNQFHIGDASAVEIKFSNGSQD--------ILLPNVIIMGAEPDLFETNSSNISLR 168
Db        121 SYSLLDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLR 177

Qy        169 ---NNYMPSNHGFGSIAIVTFSPEYSFRFNDN----------SMNEFIQDPALTLMHELI 215
Db        178 VDNKNYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELI 236

Qy        216 HSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275
Db        237 HVLHGLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKT 295

Qy        276 LADYKKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-L 331
Db        296 LNDYKAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIM 355

Qy        332 YSFTEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388
Db        356 YGFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQN 415

Qy        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGEL 434
Db        416 MRVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDL 475

Qy        435 FFVASENSYNDDNINTPKEIDDTVTSNN-----NYENDLDQVILNFNSESAPGLSDEKLN 489
Db        476 TFIAEKNSFSEEPFQ-----DEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTT 530

Qy        490 LTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQP 549
Db        531 PVTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINST 590

Qy        550 KIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLA 609
Db        591 KIYSYFPS-VISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPA 649

Qy        610 LNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINN 669
Db        650 LNIVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAE---SSTQKEKIIKTIDN 706

Qy        670 ALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLE 729
Db        707 FLEKRYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGP 766

Qy        730 EKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLM-KLINEVKINKLREYDEN 788
Db        767 DKEQIAD--EINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAK-KQLLEFDTQ 823
```

```
Qy        789 VKTYLLNYIIQHGSILGESQ-QELNSMVTDTLNNSIPFKLSSYTD---------DKILIS 838
Db        824 SKNILMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL-- 881

Qy        839 YFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSE 897
Db        882 ---------KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSE 932

Qy        898 VNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNV--NNEYTIINCMRDNN----SGWK 951
Db        933 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 992

Qy        952 VSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYI-NKWIFVTITNDRLGDSKLYINGN 1010
Db        993 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 1051

Qy       1011 LIDQKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069
Db       1052 LMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSI 1111

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLAN---------- 1119
Db       1112 TFLRDFWGNPLRYDTEYYLIPVASSS------KD--VQLKNITDYMYLTNAPSYTNGKLN 1163

Qy       1120 ----RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLFPLYADTATTNKEK 1173
Db       1164 IYYRRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSY----------------NNNEH 1207

Qy       1174 TIKISSSGNRFNQVVVMNSVGNNC---------------------TMNFKNNNGNNIGL 1211
Db       1208 IVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGL 1267

Qy       1212 LGF---------KADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGW 1249
Db       1268 VGTHNGQIGNDPNRDILIASNWYFNHLKD--KILGCDWYFVPTDEGW 1312
```

RESULT 11
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 29.0%; Score 1895; DB 2; Length 367;
  Best Local Similarity 99.7%; Pred. No. 1.1e-70;
  Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 735
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 795
Db        121 NKYDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 975
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
Db        361 NYGNANG 367
```

RESULT 12
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Krieglst
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type D
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 28.8%; Score 1883; DB 2; Length 1276;
  Best Local Similarity 34.4%; Pred. No. 1.8e-69;
  Matches 456; Conservative 250; Mismatches 487; Indels 134; Gaps 41;

```
Qy          4 INSFNYNDPVNDRTILYIK---------PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQD 54
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FHPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDNQFHIGDAS---AVEIKFSNGS---QDILLPNVIIMGAEPDLFETNSSNISLR 168
Db        118 GDSSTPEDTFDFTRHTTNIAVE-KFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQG 176

Qy        169 NNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLH 219
Db        177 QQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLH 236

Qy        220 GLYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLA 277
Db        237 QLYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALG 295

Qy        278 DYKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLY 332
Db        296 HYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLT 355

Qy        333 S-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNA 389
Db        356 NVMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNI 415

Qy        390 NLNPRIITPITGRGLVKKIIRFCKNIVSVKGIR-KSICIEINNGELFFVASENSYNDDNI 448
Db        416 ERNP-ALQKLSSESVVDLFTKVCLRL--TKNSRDDSTCIKVKNNRLPYVADKDSISQEIF 472

Qy        449 NTPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPK 500
Db        473 ENKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVF--- 529
```

```
Qy        501 YDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFI 560
              ||     || :: |: || ::||::||:     |: ||:|:: ||     |||||  |
Db        530 YD-----DITKY-VDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPS-LA 582

Qy        561 NNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGN 620
              ||| ||| ||::|  :|: |||   :| |:|||:|:|:::|||| |||||  | :||
Db        583 EKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGN 642

Qy        621 FKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKE 680
              |  |    |  ||| ||  || : |||   |  |   : |:|| | | |::| ::||:
Db        643 FNQAFATAGVAFLLEGFPEFTIPALGVFT---FYSSIQEREKIIKTIENCLEQRVKRWKD 699

Qy        681 VYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDI 740
              | ::||||:::| ||||    ||| :|  | :|||  |: :|  |:  :|  :  |  :
Db        700 SYQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENI--KSQV 757

Qy        741 KQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQH 800
              : ::| |: |:| |||||::|: | |::|| | :    |::| ::|   || |:| | |
Db        758 ENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSH 817

Qy        801 GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 860
              ||      | : | ::  |::|| : |||:: :|    |::|  |  | :|::: | :
Db        818 NIILVGEVDRLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKN 877

Qy        861 KYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS----EVNISQNDYIIYDNKYKNFS 916
              ||||||:: : :  :|       | |     |||     :: :: |: |:|   |:| |
Db        878 ALVDTSGYNAEVRVGDNVQLNTIYTNDF-----KLSSSGDKIIVNLNNNILYSAIYENSS 932

Qy        917 ISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQK---L 973
              :|||::|       : | :|||||||  : : |||||: : :   | | |||   :|:|   |
Db        933 VSFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQD---VNRKYKSL 985

Qy        974 AFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI 1033
              |:|  :   : | ||| ||||||: :|  |||||| |   : | :|  : :   |:| |
Db        986 IFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGI 1045

Qy       1034 -VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVL 1092
               |    : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|
Db       1046 DENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGNPLKFDTEYYIIN-- 1103

Qy       1093 KPNNFIDRRKDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSS------TNDN 1140
              :|:|||             | :|:       ::||:|  : |: |:: :         ||
Db       1104 --DNYIDRYIAP-------ESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDN 1154

Qy       1141 LVRK---NDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVV--VMNSVGN 1195
              ::     | : |:  :      ::       :|    :|: |:   :   :  : | |
Db       1155 IILHMLYNSRKYM-IIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSK 1213

Qy       1196 N--CTMNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDHTNSNGCFWNFIS 1244
              |  |:  | :   | : |          || | | || | | | :    |   || |||
Db       1214 NKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFIS 1269

Qy       1245 EEHGWQE 1251
                : || |
Db       1270 RDPGWVE 1276
```

RESULT 13
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membra
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

Query Match 28.5%; Score 1863; DB 2; Length 1285;
Best Local Similarity 34.5%; Pred. No. 1.2e-68;
Matches 441; Conservative 238; Mismatches 456; Indels 144; Gaps 33;

```
Qy          4 INSFNYNDPVNDRTILYIK---------PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQD 54
              :  |||:|||||  |||::          |       |:|  | :|||:||||    | |
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FHPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114
               ||          |||||:|| :||:|| ||| : |:| |||     |  |:  |   :|::
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDNQFHIGDAS---AVEIKFSNGS---QDILLPNVIIMGAEPDLFETNSSNISLR 168
              |: :||:: |      :   ||| || |||    :|:  |:|:|  |  |:: :  :|
Db        118 GDSSTPEDTFDFTRHTTNIAVE-KFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQG 176

Qy        169 NNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLH 219
               ||  |||:::|:  :||:    |:| : |:          |  || : ||||| ||||
Db        177 QQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLH 236

Qy        220 GLYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLA 277
              |||   | :   |  : :    :  | |:  ||  ||||:|: ||   :   :    |
Db        237 QLYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGSDVEIIPQIERLQLREKALG 295

Qy        278 DYKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLY 332
              || || :|: :  :  |       :: || :|  ||  ||| :| : |||:||| ::  |
Db        296 HYKDIAKRLNNINKTIPSSWSSNIDKYKKIFSEKYNFDKDNTGNFLVNIDKFNSLYSDLT 355

Qy        333 S-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNA 389
              :  :|   :::: || |  |  ::      :|:|:|:|| |  |:|:      |   |||
Db        356 NVMSEVIYSSQYNVKNRTHYFSKHYLPVFANILDDNIYTIINGFNLTTKGFNIENSGQNI 415

Qy        390 NLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNIN 449
              ||  :  ::   :|    : |  :          | ||:: |  | :|| ::| : :
Db        416 ERNP-ALQKLSSESVVDLFTKVCLRLTR-NSRDDSTCIQVKNNTLPYVADKDSISQEIFE 473

Qy        450 TPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPKY 501
              :    |:|   | :   ||: ||:      | |:     | :: | ||:  : : :   |
Db        474 SQIITDETNVENYSDNFSLDESILDAKVPTNPEAVDPLLPNVNMEPLNVPGEEEVF---Y 530

Qy        502 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 561
              |     || : ||: || ::||:|||:      |: ||:|:: ||    |||||  |
Db        531 D-----DITK-DVDYLNSYYYLEAQKLSNNVENITLTTSVEEALGYSNKIYTFLPS-LAE 583

Qy        562 NVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 621
              ||| ||| ||::|  :|: ||||   :| |:|||:|:| ::|||| |||||  | :|||
Db        584 KVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSAIIPYIGPALNIGNSALRGNF 643

Qy        622 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 681
              | |     |  ||| ||  || : |||   |  |   : |:|| | | |::| ::||:
Db        644 KQAFATAGVAFLLEGFPEFTIPALGVFT---FYSSIQEREKIIKTIENCLEQRVKRWKDS 700

Qy        682 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIK 741
              | ::||||:::| ||||    ||| :|  | :|||  |: :|  |:  :|  :  |  ::
Db        701 YQWMVSNWLSRITTQFNHISYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENI--KSQVE 758

Qy        742 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHG 801
               ::| |: |:| |||||::|: | |::|| | :    |::|  ::|   || |:| | |
Db        759 NLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHN 818

Qy        802 SILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 861
              ||       | : | ::  |:||| : |||:: :|    |::|  |  | :|::: | :
Db        819 IILVGEVDRLKAKVNESFENTIPFNIFSYTNNSLLKDMINEYFNSINDSKILSLQNKKNT 878

Qy        862 YVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYIIYDNKYKNFSISF 919
```

```
Db        879 LMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSSGDDRGKIIVTQNENIVYNAMYESFSISF 938

Qy        920 WVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGN 979

Db        939 WIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNENSEQDINFSYDI 993

Qy        980 ANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYT 1039

Db        994 SKNAAGY-NKWFFVTITTNMMGNMMIYINGKLIDTIKVKELTGINFSKTITFQMNKIPNT 1052

Qy       1040 RYI---------GIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLN 1090

Db       1053 GLITSDSDNINMWIRDFYIFAKELDDKDINILFNSLQYTNVVKDYWGNDLRYDKEYYMIN 1112

Qy       1091 VLKPNNFIDR---RKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQ 1147

Db       1113 V----NYMNRYMSKKGNGIVFNTRKNN----NDFNEGYKIIIKRI-RGNTNDTRVRGENV 1163

Qy       1148 VYIN----------------------------------------FVASKTHLFPLYA-- 1164

Db       1164 LYFNTTIDNKQYSLGMYKPSRNLGTDLVPLGALDQPMDEIRKYGSFIIQPCNTFDYYASQ 1223

Qy       1165 ----DTATTNKEKTIKISS 1179

Db       1224 LFLSSNATTNRLGILSIGS 1242


RESULT 14
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

  Query Match           27.9%;  Score 1822;  DB 2;  Length 1291;
  Best Local Similarity  33.8%;  Pred. No. 5.6e-67;
  Matches  451;  Conservative  256;  Mismatches  487;  Indels  142;  Gaps   43;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         57 PP--TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114

Db         61 PPRVTSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFP 117

Qy        115 GNDNTPDNQFHIG-DASAVEIKFSNGSQDI----LLPNVIIMGAEPDLFETNSSNISLRN 169

Db        118 GNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTN 177

Qy        170 NYMPSNHGFGSIAIVTFSPE----YSFRFND-----NSMNEFIQDPALTLMHELIHSLHG 220

Db        178 NTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHN 237

Qy        221 LYGAKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNLLAD 278

Db        238 LYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDY 296

Qy        279 YKKIASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS 333

Db        297 YRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQ 356

Qy        334 -FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNAN 390

Db        357 IFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLS 416

Qy        391 LNPRIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASENSYNDD 446

Db        417 RNP-ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTD 475

Qy        447 -----NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSDEKLNLT 491

Db        476 IFLRKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDSESEILP 527

Qy        492 IQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKI 551

Db        528 GENQVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKV 578

Qy        552 YTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALN 611

Db        579 YTYFPT-LANKVNAGVQQGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637

Qy        612 IGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNAL 671

Db        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIKTIDNCL 694

Qy        672 KERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEK 731

Db        695 EQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDK 754

Qy        732 NELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKT 791

Db        755 ENI--KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKA 812

Qy        792 YLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSS 851

Db        813 KLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSK 872

Qy        852 VLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYIIYD 909

Db        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYN 932

Qy        910 NKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGI 969

Db        933 SMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDS 987

Qy        970 NQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNI 1029

Db        988 EQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTI 1046

Qy       1030 LFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Db       1047 TFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDL 1106

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Db       1107 RYNKEYYMVNIDYLNRYM--YANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RGNTNDT 1160

Qy       1141 LVRKNDQVYINFVASKT--HLF----PLYADTATTN-------KEKTIKISSSGNRFNQV 1187

Db       1161 RVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDIND--NIIFQI 1218

Qy       1188 VVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVASTWY-------------YT 1227

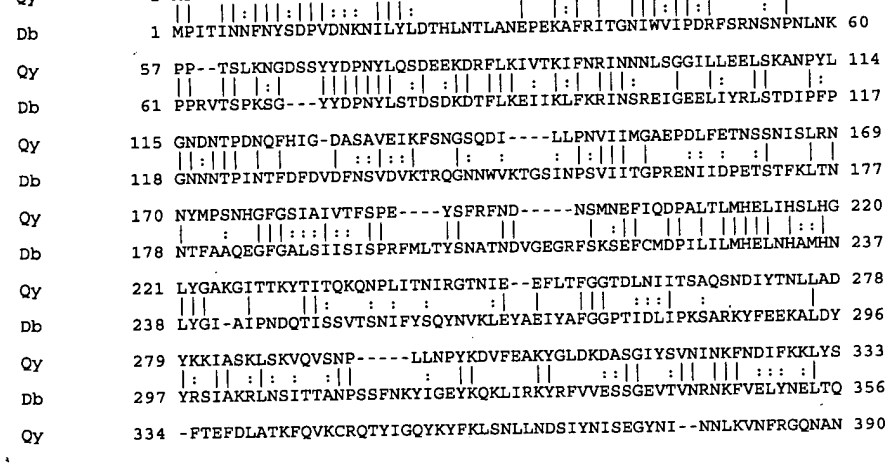
Db       1219 QPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYA 1274

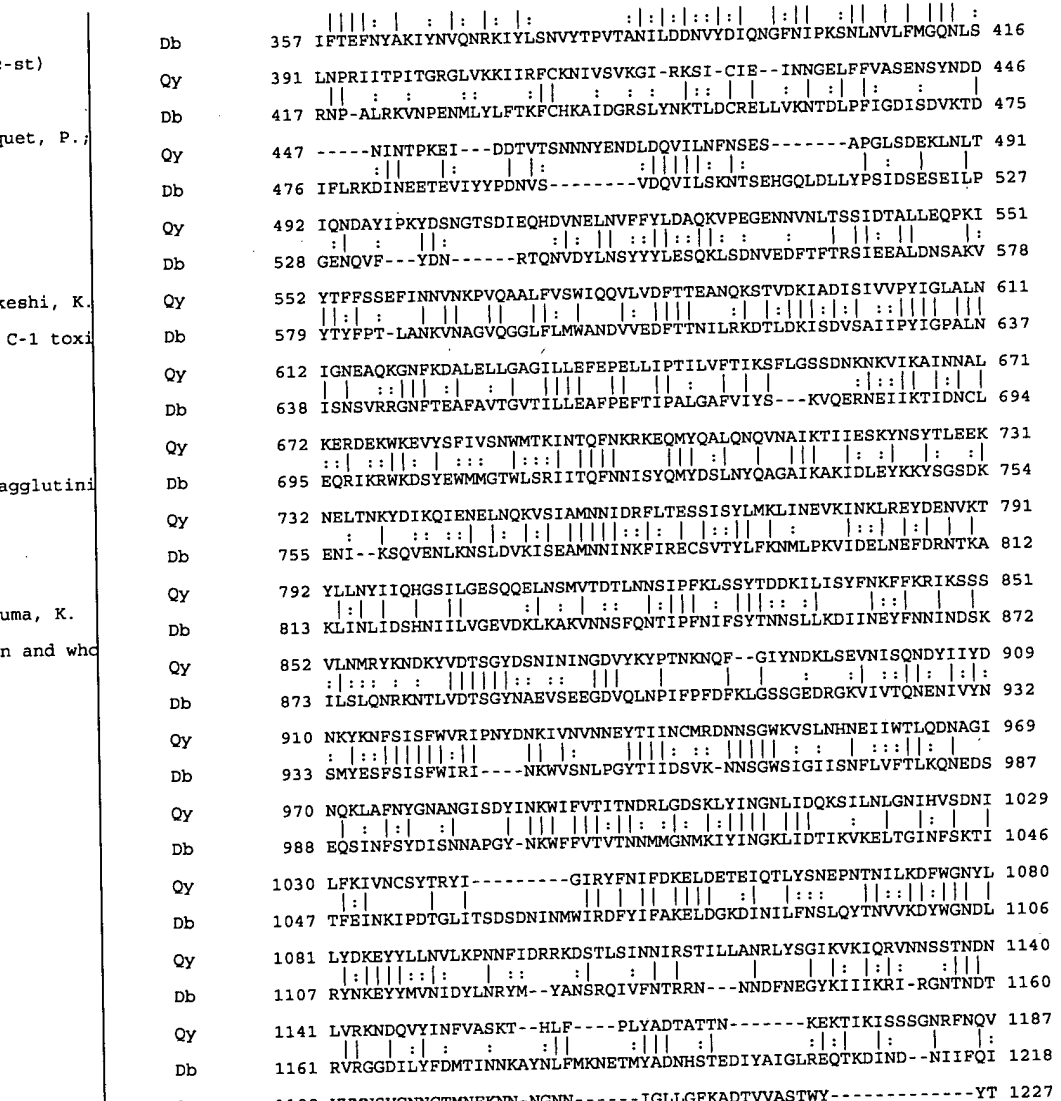
Qy       1228 HMRDHTNSNGCFWNFI 1243

Db       1275 SLLESTSTH---WGFV 1287
```

RESULT 15
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 27.9%; Score 1822; DB 2; Length 1291;
Best Local Similarity 33.8%; Pred. No. 5.6e-67;
Matches 451; Conservative 256; Mismatches 487; Indels 142; Gaps 43;





Search completed: September 10, 2004, 03:44:07
Job time : 27.1185 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 15.2761 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 6541
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6487.5	99.2	1250	1	BXE_CLOBO	Q00496 clostridium
2	6339.5	96.9	1250	1	BXE_CLOBU	P30995 clostridium
3	4056.5	62.0	1274	1	BXF_CLOBO	P30996 clostridium
4	2339	35.8	1295	1	BXA1_CLOBO	P10845 clostridium
5	2285	34.9	1295	1	BXA2_CLOBO	Q45894 clostridium
6	2254.5	34.5	1296	1	BXG_CLOBO	Q60393 clostridium
7	2203.5	33.7	1290	1	BXB_CLOBO	P10844 clostridium
8	2151	32.9	1314	1	TETX_CLOTE	P04958 clostridium
9	1883	28.8	1276	1	BXD_CLOBO	P19321 clostridium
10	1820.5	27.8	1290	1	BXC1_CLOBO	P18640 clostridium
11	761.5	11.6	1196	1	BXCN_CLOBO	P46081 clostridium
12	747.5	11.4	1162	1	BXEN_CLOBO	P46082 clostridium
13	734.5	11.2	1162	1	BXEN_CLOBU	Q06366 clostridium
14	293	4.5	2136	1	YCF2_MARPO	P09975 marchantia
15	277	4.2	1956	1	ATX1_PLAFA	Q04956 plasmodium
16	275	4.2	2867	1	RBP2_PLAVB	Q00799 plasmodium
17	259.5	4.0	2710	1	TOXA_CLODI	P16154 clostridium
18	257	3.9	2894	1	YD96_METJA	Q58791 methanococc
19	256	3.9	1024	1	RPOB_PLAFA	P21421 plasmodium
20	256	3.9	2339	1	RPC1_PLAFA	P27625 plasmodium
21	244	3.7	1738	1	YCF1_EPIVI	Q00383 epifagus vi
22	241.5	3.7	964	1	DPOL_CBEPV	P30319 choristoneu
23	236	3.6	1892	1	Y835_RICCN	Q92hd6 rickettsia
24	235.5	3.6	1225	1	Y309_MYCGE	P47551 mycoplasma
25	234.5	3.6	2366	1	TOXB_CLODI	P18177 clostridium
26	232.5	3.6	1169	1	EX5B_BORBU	O51578 borrelia bu
27	230.5	3.5	968	1	Y080_BUCAP	Q8ka32 buchnera ap
28	229	3.5	760	1	ETF2_MSEPV	Q9yw29 melanoplus
29	228.5	3.5	2869	1	RBP1_PLAVB	Q00798 plasmodium
30	227	3.5	2971	1	YCX9_CHLRE	Q32065 chlamydomon
31	224	3.4	1901	1	YCF1_TOBAC	P12222 nicotiana t
32	222	3.4	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
33	221.5	3.4	1407	1	CYAA_DICDI	Q03100 dictyosteli
34	221	3.4	1271	1	Y338_MYCGE	P47580 mycoplasma
35	221	3.4	1682	1	MSP1_PLAF3	P19598 plasmodium
36	220.5	3.4	1167	1	CAGA_HELPJ	Q9zlt1 helicobacte
37	219.5	3.4	1157	1	N133_YEAST	P36161 saccharomyc
38	218.5	3.3	971	1	Y228_BORBU	O51246 borrelia bu
39	218.5	3.3	1177	1	Y307_MYCGE	P47549 mycoplasma
40	218.5	3.3	3418	1	BRC2_HUMAN	P51587 homo sapien
41	218	3.3	1111	1	KIP1_YEAST	P28742 saccharomyc
42	218	3.3	1630	1	MSP1_PLAFK	P04932 plasmodium
43	218	3.3	1639	1	MSP1_PLAFW	P04933 plasmodium
44	217.5	3.3	1113	1	Y140_MYCPN	P75033 mycoplasma
45	217.5	3.3	2452	1	RPB1_PLAFD	P14248 plasmodium

ALIGNMENTS

RESULT 1
BXE_CLOBO
ID BXE_CLOBO STANDARD; PRT; 1250 AA.
AC Q00496;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beluga;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinal neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [4]
RP SEQUENCE OF 1-13.
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [5]
RP SEQUENCE OF 419-426.
RX MEDLINE=90344918; PubMed=2116911;
RA Gimenez J.A., Dasgupta B.R.;
RT "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RT neurotoxin.";
RL Biochimie 72:213-217(1990).
RN [6]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,

```
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match            99.2%;  Score 6487.5;  DB 1;  Length 1250;
  Best Local Similarity  99.4%;  Pred. No. 3.8e-255;
  Matches 1243;  Conservative    2;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      2 PKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSL 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 PKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSL 60

Qy     62 KNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPD 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPD 120

Qy    122 NQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 NQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHRFGSI 180

Qy    182 AIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 241
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AIVTFSPEYSFRFNDNCMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 240

Qy    242 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 300

Qy    302 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 361
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    301 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKLS 360

Qy    362 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 420

Qy    422 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 480

Qy    482 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 540

Qy    542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 600

Qy    602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 660

Qy    662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIES 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIES 720

Qy    722 KYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    721 KYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINK 780

Qy    782 LREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFN 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 LREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFN 840

Qy    842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 900

Qy    902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 960

Qy    962 TLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLG 1021
          | :|| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLG 1020

Qy   1022 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLL 1081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLL 1080
```

```
Qy       1082 YDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1141
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 YDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1140

Qy       1142 VRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNF 1201
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1141 VRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-NCTMNF 1199

Qy       1202 KNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 KNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1250


RESULT 2
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             96.9%;  Score 6339.5;  DB 1;  Length 1250;
  Best Local Similarity   97.3%;  Pred. No. 3.7e-249;
  Matches 1217;  Conservative   15;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          2 PKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSL 61
              | |||||||||||:|||||||||||:|||||||||||||||||||| |||| |||||
Db          1 PTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTSL 60

Qy         62 KNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPD 121
              ||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 KNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTPD 120

Qy        122 NQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 181
              | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 GDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSI 180

Qy        182 AIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 241
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPLI 240

Qy        242 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKD 300

Qy        302 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKLS 360

Qy        362 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI 420

Qy        422 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAP 480

Qy        482 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSI 540

Qy        542 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISI 601
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        541 DTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISI 600
```

```
Qy       602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 661
Db       601 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKN 660

Qy       662 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIES 721
Db       661 KVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIES 720

Qy       722 KYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 781
Db       721 KYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINK 780

Qy       782 LREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFN 841
Db       781 LREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFN 840

Qy       842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
Db       841 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 900

Qy       902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
Db       901 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 960

Qy       962 TLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLG 1021
Db       961 TLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLG 1020

Qy      1022 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLL 1081
Db      1021 NIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLL 1080

Qy      1082 YDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1141
Db      1081 YDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL 1140

Qy      1142 VRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNF 1201
Db      1141 VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVG-NCTMNF 1199

Qy      1202 KNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
Db      1200 KNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1250
```

```
RESULT 3
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             62.0%;  Score 4056.5;  DB 1;  Length 1274;
```

```
  Best Local Similarity   61.0%;  Pred. No. 6.3e-157;
  Matches  781;  Conservative  221;  Mismatches  242;  Indels   37;  Gaps   16;

Qy          1 MP-KINSFNYNDPVNDRTILYIK---PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  |||||||||||| ||||::       :::||:| ||:|:||||||| ||| | ||
Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTNPSDFD 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
              || ||||| |:|||||||| :| ||||:||    |:| |||:| :| :||:|:| | |||||
Db         61 PPASLKNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYLGN 120

Qy        117 DNTPDNQFH-IGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRN------ 169
              |:|| ::|  :   ::| || |   :  :| |::::|| ||:||  |    :|
Db        121 DHTPIDEFSPVTRTTSVNIKLSTNVESSMLLNLLVLGAGPDIFE--SCCYPVRKLIDPDV 178

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFND------NSMNEFIQDPALTLMHELIHSLHGLYG 223
              | |||:||||| |||||||| : |||      :|   || |||::| |||||:||||||
Db        179 VYDPSNYGFGSINIVTFSPEYEYTFNDISGGHNSSTESFIADPAISLAHELIHALHGLYG 238

Qy        224 AKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIA 283
              |:|:| : ||  || ||:   :  :|||||||| ||||||||     || ||||:|:|||
Db        239 ARGVTYEETIEVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIA 298

Qy        284 SKLSKVQVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLAT 341
              ::||:|  : |   :| ||| |: ||||||:| | |:|| |||||:|:||||||| |||
Db        299 TRLSEVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTESDLAN 358

Qy        342 KFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITG 401
              ||:|||| ||  :|:: |: |||:| || :|||:|| || || |||:  |||:|| |
Db        359 KFKVKCRNTYFIKYEFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQSIKLNPKIIDSIPD 418

Qy        402 RGLVKKIIRFCKNIVSVKGIR--KSICIEINNGELFFVASENSYNDDNINTPKEIDDTVT 459
              :|||:||::|||:::  || :    :|| :|| |||||||:|||:::||||||||||
Db        419 KGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNSELFFVASESSYNENDINTPKEIDDTTN 478

Qy        460 SNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNV 519
              |||| |:||:|||::||:: | :|:  ||  :|:::|:|:|||||||:||::|| : ||
Db        479 LNNNYRNNLDEVILDYNSQTIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYDVVDFNV 538

Qy        520 FFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQV 579
              |||| |||||||| |::||||||||||||: |   |||||||: :|||| |||| || :|
Db        539 FFYLHAQKVPEGETNISLTSSIDTALLEESK-DIFFSSEFIDTINKPVNAALFIDWISKV 597

Qy        580 LVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPE 639
              : |||||| ||||||||||||::|||:||||||  ||:||||::| |||| |||||| ||
Db        598 IRDFTTEATQKSTVDKIADISLIVPYVGLALNIIIEAEKGNFEEAFELLGVGILLEFVPE 657

Qy        640 LLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNK 699
              | || ||||||||:: | :|||| ||||||:| ||: ||||:||:|||||:|:|||||||
Db        658 LTIPVILVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRINTQFNK 717

Qy        700 RKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNID 759
              |||||||||||||:|||| || |||:|| :||| | ::|:|  || |||:|||:|| ||:
Db        718 RKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRLESEYNINNIEEELNKKVSLAMKNIE 777

Qy        760 RFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTL 819
              ||:|||||||||||||| |: ||::|| :||: |||||: | |||||   ||: :|| ||
Db        778 RFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLLNYILDHRSILGEQTNELSDLVTSTL 837

Qy        820 NNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVY 879
              |:||||:|||||:||||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:||
Db        838 NSSIPFELSSYTNDKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVY 897

Qy        880 KYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTI 939
              | ||:|||||||:|||||:||: |||:::|:|||||||||| :  | :| | ||||
Db        898 IYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTI 956

Qy        940 INCMRDNNSGWKVSL---NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTIT 996
              |||| :||||||:||      |||||||| :|  : | | |    | ||:|||||||||||
Db        957 INCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTIT 1016

Qy        997 NDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDE 1056
              |:|||:|::||||||| :||| |||:|||||||||||| |    |:|||| :|: |||:

Db       1017 NNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDK 1076

Qy       1057 TEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIR---- 1112
              |||:|||||||: :|||::|||||||:|:||| |:|: : :|     |:||  |
Db       1077 TEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYI-TLNSGILNINQQRGVTE 1135

Qy       1113 STILLANRLYSGIKVKIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTN 1170
                :: |  :|| |::| |::    : |  || |||||  ||| |  :   : |||||  :
Db       1136 GSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSE 1193

Qy       1171 KEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMR 1230
              ||| |: |:  :   |::||:|:||||||||:||||:||||||| :: :|||:||| ::|
Db       1194 KEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIR 1253

Qy       1231 DHTNSNGCFWNFISEEHGWQE 1251
                :|:|||||: ||:|:||:|
Db       1254 RNTSSNGCFWSSISKENGWKE 1274


RESULT 4
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
```

```
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
```

```
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match           35.8%;  Score 2339;  DB 1;  Length 1295;
  Best Local Similarity   40.0%;  Pred. No. 1.5e-87;
  Matches  536;  Conservative  248;  Mismatches  420;  Indels  136;  Gaps   44;

Qy          2 PKIN-SFNYNDPVNDRTILYIK-PG-GCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFH 56
              | :|  ||| ||||   | ||| |  |  :  |:| |   ||:||||:   | |:  | :
Db          1 PFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLN 59

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
              ||   |   ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        117 DNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPS 174
              :| | :  :  | : : :   :||       |::|:|   |:  ||  |    : |   :
Db        120 -STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLN---LT 175

Qy        175 NHGFGSIAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAK 225
              :|:||   : |||:::| | ::            :|  |||:|| |||||: | ||
Db        176 RNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLY--- 232

Qy        226 GITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADY 279
              ||        | |   ||    : |  :  ||  |||| |   | | | |:       :
Db        233 GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKF 289

Qy        280 KKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 335
              | ||| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|
Db        290 KDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT 349

Qy        336 EFDLATKFQVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 392
              | :    |:|  |:||:   |  ||: |::    | | :|:|:   ||  || ||| :|
Db        350 EDNFVKFFKVLNRKTYLNFDKAVFKI-NIVPKVNYTIYDGFNLRNTNLAANFNGQNTEIN 408

Qy        393 PRIITPI---TGRGLVKKIIRFCKNIVSVKGIRKS--------------ICIEINNGEL 434
                | :   ||       :  | | :: |:||  |              :||::|| :|
Db        409 NMNFTKLKNFTG------LFEFYK-LLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDL 461

Qy        435 FFVASENSYNDDNINTPKEIDDTVTSNNNYE--------NDLDQVILNFNSESAP-GLSD 485
              ||  ||::: :| :|  :||    ||: | |        : : |  | || :: |  :|
Db        462 FFSPSEDNFTND-LNKGEEI----TSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISI 516

Qy        486 EKLNLTIQND-AYIPKYD--SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSID 542
              | |:  |       :|  :   ||    :::::::  :| || ||:   |:: : ||:|::
Db        517 ENLSSDIIGQLELMPNIERFPNG----KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVN 572

Qy        543 TALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIV 602
               |||   ::||||||:::  |||  :||:|: |::|:: ||| | :: || ||||||:|:
Db        573 EALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITII 632

Qy        603 VPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNK 662
              :|||| |||||   | :|  ||   || ||||| ||: || :  | : |::    ||
Db        633 IPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVL 689

Qy        663 VIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESK 722
               :: |:||| :|:||| ||| :||:||: |:||| :  :::| :||:||  | | ||  :
Db        690 TVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ 749

Qy        723 YNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKL 782
              || || |||| :   ::|  : ::||: :: || ||::|| : |:|||  :    : :|
Db        750 YNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRL 807

Qy        783 REYDENVKTYLLNYIIQH-GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFN 841
               ::| ::|  || ||  : |:::|:  : |   | :||:  |||:|| | |:: |:| |
Db        808 EDFDASLKDALLKYIYDNRGTLIGQVDR-LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFT 866

Qy        842 KFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNIS 901
              :: | | ::|:||:||:::  :| | | | |||    |   | :|||  ::| : |:: :
Db        867 EYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVI 926

Qy        902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIW 961
              : |:|:: |:||| |||:||| | | | ::|||||||||| :|||||||||: ||||
Db        927 LKNAIVYNSMYENFSTSFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSLNYGEIIW 984

Qy        962 TLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLG 1021
              ||||   | |:: | |      ||||||:||||||||:|| :||:|||| |||| | |||
Db        985 TLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG 1044

Qy       1022 NIHVSDNILFKIVNCSYT-RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||| |:||:||:   |  | ||| |:|||:|||||:| ||: || |: |: |||||||:||
Db       1045 NIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYL 1104

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNI----------------RSTILLANRLYSG 1124
              ||| ||:||:  || ::|       :||:                  : | | : || |
Db       1105 QYDKPYYMLNLYDPNKYVD--------VNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy       1125 IKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEK---TIKISSSG 1181
              |  |::   |   ||:|| ||:|||| |  |   : |  : :    ||    ::|   |
Db       1157 TKFIIKKY-ASGNKDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVEKILSALEIPDVG 1214

Qy       1182 NRFNQVVVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVVASTWYYTHMRDH 1232
              |  :|||| |      : | | || ::|||||:|| :||        :||| ||   :
Db       1215 N-LSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1273

Qy       1233 TNSNGCFWNFISEEHGWQEK 1252
              : : || | ||  : || |:
Db       1274 SRTLGCSWEFIPVDDGWGER 1293


RESULT 5
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
```

```
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             34.9%;  Score 2285;  DB 1;  Length 1295;
  Best Local Similarity   38.6%;  Pred. No. 2.3e-85;
  Matches  522;  Conservative  240;  Mismatches  429;  Indels  160;  Gaps   38;

Qy          2 PKIN-SFNYNDPVNDRTILYIK--PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQ--DFH 56
              | :|  ||| ||||   | |||    |  :  |:| |   ||:||||:   | |:  | :
Db          1 PFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLN 59

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
              ||    |    ||||  || :| ||| :|| |||:| || :   | :||  : :  |: |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        117 DNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISLRNNYMPS 174
               :| | :  : | : :  :||       |::|:|   |:  ||  |    : |    :
Db        120 -STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHDVLN---LT 175

Qy        175 NHGFGSIAIVTFSPEYSFRFNDN---------SMNEFIQDPALTLMHELIHSLHGLYGAK 225
              :|:||   : |||:::| | ::           :|  |||:|| |||||: | ||
Db        176 RNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAEHRLY--- 232

Qy        226 GITTKYTITQKQNPLITN----IRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLADY 279
              ||         | |   ||    : |  :  ||  |||| |   | | | |:        :
Db        233 GIAINPNRVFKVN---TNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKF 289

Qy        280 KKIASKLSKVQV---SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 335
              | :|| |:| :    :   |   |:||: || | :| || :||:  ||: ::| |   :|
Db        290 KDVASTLNKAKSIIGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT 349

Qy        336 EFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
              | :    |:|  |:||:   |     |:: |  | | :|:|:   ||  || |||  :|

Db        350 EDNFVNFFKVINRKTYLNFDKAVFRINIVPDENYTIKDGFNLKGANLSTNFNGQNTEINS 409

Qy        394 RIITPI---TGRGLVKKIIRFCKNIVSVKGI--------------RKSICIEINNGELF 435
              |  | :   ||      :  | | :: |:||                 :||::|| :||
Db        410 RNFTRLKNFTG------LFEFYK-LLCVRGIIPFKTKSLDEGYNKALNDLCIKVNNWDLF 462

Qy        436 FVASENSYNDDNINTPKEID--DTVTSNNNYE--------NDLDQVILNFNSESAPGLSD 485
              |  ||::: :|       :|  : :|:: | |         : :  |  | |: :: |
Db        463 FSPSEDNFTND-------LDKVEEITADTNIEAAEENISLDLIQQYYLTFDFDNEPE--- 512

Qy        486 EKLNLTIQNDAYIPKYDSNGTSDI---------------EQHDVNELNVFFYLDAQKVP 529
                |::|:          | :|||                :::::::  :| || ||:
Db        513 ---NISIE----------NLSSDIIGQLEPMPNIERFPNGKKYELDKYTMFHYLRAQEFE 559

Qy        530 EGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQ 589
               |:: : ||:| : |||:    ||||||:::  :|| |:| :|::| :::: ||| | |:
Db        560 HGDSRIILTNSAEEALLKPNVAYTFFSSKYVKKINKAVEAFMFLNWAEELVYDFTDETNE 619

Qy        590 KSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFT 649
              :|:|||||:|:|||| ||||||   || | :|:   |   :||| ||  :|    |
Db        620 VTTMDKIADITIIVPYIGPALNIGNMLSKGEFVEAIIFTGVVAMLEFIPEYALPVFGTFA 679

Qy        650 IKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQ 709
              | |::    ||   :: ||||| :|:||| ||| : |:||: |:||| :  :|:| :||:
Db        680 IVSYIA---NKVLTVQTINNALSKRNEKWDEVYKYTVTNWLAKVNTQIDLIREKMKKALE 736

Qy        710 NQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISY 769
              ||  | | ||  :|| || |||| :   ::|  : ::||: :: || ||::|| : |:||
Db        737 NQAEATKAIINYQYNQYTEEEKNNI--NFNIDDLSSKLNESINSAMININKFLDQCSVSY 794

Qy        770 LMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSS 829
              ||  :   : :|:::| :|:  || ||  :   |       |   | :||:  |||:||
Db        795 LMNSMIPYAVKRLKDFDASVRDVLLKYIYDNRGTLVLQVDRLKDEVNNTLSADIPFQLSK 854

Qy        830 YTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFG 889
              | |:| |:|  | :: | | ::|:|:: || |  :| | | : |||   ||    :|||
Db        855 YVDNKKLLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIK 914

Qy        890 IYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSG 949
              : | : | : :  : |:|:: |:||| |||::|| | :|| |:||||||||: :||||
Db        915 LINLESSTIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKI-NLNNEYTIINCI-ENNSG 972

Qy        950 WKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYING 1009
              ||||||: |||||||||    |:: | |    |||||:|||||||:||  ||:||||
Db        973 WKVSLNYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYING 1032

Qy       1010 NLIDQKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYSNEPN 1068
              ||||| | |||||| |: |:||:  |     ||| |:|||:|||||:| ||: || :: |
Db       1033 RLIDQKPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSN 1092

Qy       1069 TNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNI----------------R 1112
              : |||||||||| ||| ||:||:  || ::|        :|||
Db       1093 SGILKDFWGNYLQYDKPYYMLNLFDPNKYVD--------VNNIGIRGYMYLKGPRGSVVT 1144

Qy       1113 STILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKE 1172
              : | | : || | |  |::    |   ||:|| ||:|||| |  |   : |  : :    |
Db       1145 TNIYLNSTLYEGTKFIIKKY-ASGNEDNIVRNNDRVYINVVV-KNKEYRLATNASQAGVE 1202

Qy       1173 K---TIKISSSGNRFNQVVVMNS-----VGNNCTMNFKNNNGNNIGLLGF----KADTVV 1220
              |    ::|   ||  :|||| |     : | | || ::||||:|| :||       :|
Db       1203 KILSALEIPDVGN-LSQVVVMKSKDDQGIRNKCKMNLQDNNGNDIGFIGFHLYDNIAKLV 1261

Qy       1221 ASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1251
              || ||   :   : : || | ||  : || |
Db       1262 ASNWYNRQVGKASRTFGCSWEFIPVDDGWGE 1292


RESULT 6
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
```

```
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             34.5%;  Score 2254.5;  DB 1;  Length 1296;
  Best Local Similarity   38.4%;  Pred. No. 4e-84;
  Matches  511;  Conservative  252;  Mismatches  449;  Indels  117;  Gaps   34;

Qy          4 INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              |  ||||||:|:  |: ::|    |   :||:| |:  |||:|||   |  |  |:  |
Db          4 IKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 63

Qy         61 LKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNT 119
              : : |   |||| ||::| |||:||| : |:|||||:  ||  ||: :  | ||||| :|
Db         64 VFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAST 123

Qy        120 PDNQF--HIGDASAVEIKFSNGSQDI---LLPNVIIMGAEPDLFETNSSNISLRNNYMPS 174
              | ::|  :: : |  :      |::|    |: |:|| |  | | : | ::  : | : |
Db        124 PPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNGHSPI 182

Qy        175 NHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGLYGAK 225
              : |||:  :: | |      ||:   |:         :  ||||||||||||| |||||| |
Db        183 SEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIK 242

Qy        226 ----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADY 279
                  ||  ||     | :|:          ||  |||| | ::|: :   :||    | ::
Db        243 ISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKALQNF 295

Qy        280 KKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFT 335
              : ||::|   |  | |  :: || ::: ||   :| :| |||: :||: ::| | : ||
Db        296 QDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFT 355

Qy        336 EFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 392
              | :|| :: :| | :|  :|  ·|   ||:::||  :||:||   |||  | |||  :|
Db        356 ETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVN 415

Qy        393 PRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPK 452
                    |:   ||   | || ::    :   || :|| :|||:|:::|:: |
Db        416 KEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKA-- 473

Qy        453 EIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYDSNGT 506
                :|:   | ||  ||:  :||:||: :  |   | :|    | :|  || |
Db        474 ---ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQ-- 528

Qy        507 SDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKP 566
              | :::  |:  ::| || ||  |    |: ||:|:: ||    |:|||||:  :   |
Db        529 SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTV 588

Qy        567 VQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALE 626
              | |:|||:|:: |: |||:|: ||||:|||::|:|:|||::|||| |||:|||  | |||:| |
Db        589 VGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFE 648

Qy        627 LLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIV 686
              : || ||:|| |||::| :  ||::|::|    ||  :|  |:||||:||:|| ::|  ||
Db        649 IGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGLIV 705

Qy        687 SNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENE 746
              | |:: :||||   ||:|| || ||  ||: ||| :|| |: |:|  :    |   |: :
Db        706 SQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFNDIDFK 763

Qy        747 LNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGE 806
              ||| :::|:|||| |: : |||||  :  : : ||:::|:|:|  || ||  :   | :
Db        764 LNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLD 823

Qy        807 SQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTS 866
                  | | |   | :|||| || || | |||  || :   | |:::|:: |:  : :|:|
Db        824 EVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 883

Qy        867 GYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNY 926
              || : :|:  ||         || : | : | :   |: :::||: : ||||:|||| | |
Db        884 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 943

Qy        927 DNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGIS 984
              :|  :   : ||||||:|:: |:|||||:  | |||||  |    :: : | |   : ||
Db        944 NNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNIS 1002

Qy        985 DYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCS-YTRYIG 1043
              ||||||  :|||||||:: :||||:|   : ||||  |: |::| ||::||:  |:::
Db       1003 DYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVW 1062

Qy       1044 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFI----D 1099
              |: |||| :||: ||: :||  : :|| ||||||| | || :||| |    | :|
Db       1063 IKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSK 1122

Qy       1100 RRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN------ 1151
```

```
                    |    |  :  :    || |::  |:: :||    |||:||: | :|:|
Db       1123 ASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISD 1182

Qy       1152 -------FVASK---THLF-------PLYADTATTNK--EKTIKISSSGNRFNQVVVMNS 1192
                  | ||   | ||       | :  |     | |||         :|  ::
Db       1183 ESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKT--------TYNCQIL--- 1231

Qy       1193 VGNNCTMNFKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDHTNS--NGCFWNF 1242
                  |  : |      ||      |:  ||        | ||   : :: |    || | |
Db       1232 ----CEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQF 1287

Qy       1243 ISEEHGWQE 1251
                  |  : || |
Db       1288 IPVDEGWTE 1296


RESULT 7
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
```

SQ SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;

Query Match 33.7%; Score 2203.5; DB 1; Length 1290;
Best Local Similarity 37.5%; Pred. No. 4.5e-82;
Matches 490; Conservative 271; Mismatches 466; Indels 81; Gaps 31;

```
Qy          4 INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||:||||||:::  |: ::|    |   :||:| |   |||||||    |  |:||:  :
Db          4 INNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNKSSG 63

Qy         61 LKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNT 119
               : | |   ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||:
Db         64 IFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLGDRRV 123

Qy        120 PDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNNYMPS 174
              |  :|:   ||    |  :   ::     :  |:|| |  | | |  : :| ::|::  |
Db        124 PLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHF-AS 182

Qy        175 NHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGLYGAK 225
              ||| |  : | |||   ||:   |:         :  |||| ||||||| |||||| |
Db        183 REGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLYGIK 242

Qy        226 GITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
               :    :  ::   : :      ||  |||| | :||| :    ||  :| ::: |  :
Db        243 -VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIVDR 301

Qy        286 LSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTEFDLA 340
              |:|| |     |  :| ||: |: ||   :|: | ||:::  |: ::| | : ||| ::|
Db        302 LNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETNIA 361

Qy        341 TKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNPRIIT 397
               :::| | :|       |: |||:: || | ||:||:  :::  :|||  :| :
Db        362 ENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQAYE 421

Qy        398 PITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT 457
              |:   |    |: ||   |||     |||:::| :|||:| :|||::|| ::  : |:
Db        422 EISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNERIEYN 475

Qy        458 VTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVN 515
              | :|  |||  ::::||: :  |   | |       :  :| |:      |::   :
Db        476 -TQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPA--IKKIFTD 532

Qy        516 ELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSW 575
              |  :| || :|   |     :::|||| | |||   |:|:||| ::|    || |:| ||  |
Db        533 ENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGW 592

Qy        576 IQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLE 635
              ::|:: ||  |||: :|:|||||||::|||||||||:|||  ||||::| |: || ||||
Db        593 VKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLE 652

Qy        636 FEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINT 695
              | |||||| :  | ::|::   |||||:|| |:||| :|:||| ::|  ||: |:: :||
Db        653 FIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNT 709

Qy        696 QFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAM 755
              ||   || ||:||  |  |:: ||: :|| |: :||: :     |   | ::||: :: |:
Db        710 QFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGINQAI 767

Qy        756 NNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMV 815
              :||: |:   |:|||| :  : : || ::|  :|  |||| ::   |  | :   | |
Db        768 DNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKV 827

Qy        816 TDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININ 875
              |   :|| || ||:| |||  |||:   | :: :||:|||::  :| ||| : : :
Db        828 NKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVY 887

Qy        876 GDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--V 933
              |     :|||| : :   |:: ::||  ||::: : :||:|||:||| | |  :   :
Db        888 DGVEL--NDKNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYI 945

Qy        934 NNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFV 993
              :|||||||||: ||||||:|:  | ||||| |  |  : : | |     ||:|||:| ||
Db        946 HNEYTIINCMK-NNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFV 1004

Qy        994 TITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDK 1052
              ||||: | ::|:|||| |      | ::  :  :  |:||:  :   |::| ::||:||:
Db       1005 TITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNT 1063

Qy       1053 ELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS------TL 1106
              || :: |:  |  :  :  |||||| |:|:||||: |    |::|  :|||       |
Db       1064 ELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTR 1123

Qy       1107 SINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLFPLYAD 1165
              |  |  |  :    || | |  |:| :|| | ||::||| | :|::|  :    : :|
Db       1124 SKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFF-NLNQEWRVYTY 1182

Qy       1166 TATTNKEKTI---KISSSGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLGFK-- 1215
                   :|: :    || |   :| : :          :| : || :  : : |||:|
Db       1183 KYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRF 1242

Qy       1216 ----------ADTVVASTWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251
                        |    | ||   :  : :    || | || :: || |
Db       1243 YESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1290
```

RESULT 8
TETX_CLOTE
ID TETX_CLOTE STANDARD; PRT; 1314 AA.
AC P04958;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE chain (Tetanus toxin chain H)].
GN TETX OR CTP60.
OS Clostridium tetani.
OG Plasmid pE88, and Plasmid 75 Kbp.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC PLASMID=75 Kbp;
RX MEDLINE=87053814; PubMed=3536478;
RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA Weller U., Hudel M., Habermann E., Niemann H.;
RT "Tetanus toxin: primary structure, expression in E. coli, and
RT homology with botulinum toxins.";
RL EMBO J. 5:2495-2502(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CN3911; PLASMID=75 Kbp;
RX MEDLINE=87040747; PubMed=3774547;
RA Fairweather N.F., Lyness V.A.;
RT "The complete nucleotide sequence of tetanus toxin.";
RL Nucleic Acids Res. 14:7809-7812(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88; PLASMID=pE88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN [4]
RP SEQUENCE OF 742-1314 FROM N.A.
RC PLASMID=75 Kbp;
RX MEDLINE=86085672; PubMed=3510187;
RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT fragment C in Escherichia coli.";
RL J. Bacteriol. 165:21-27(1986).
RN [5]

```
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
```

```
FT   STRAND      1163   1166
FT   STRAND      1173   1178
FT   TURN        1184   1185
FT   STRAND      1188   1188
FT   STRAND      1190   1190

  Query Match             32.9%;  Score 2151;  DB 1;  Length 1314;
  Best Local Similarity   35.8%;  Pred. No. 6.1e-80;
  Matches  488;  Conservative  260;  Mismatches  443;  Indels  172;  Gaps   36;

Qy          4 INSFNYNDPVNDRTILYIKPGGCQE---FYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||:| |:|||| ||: ::|  |:    :||:| |   |||:|||   || |:||:||:|
Db          4 INNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFNPPSS 63

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |  | | |||||||::| :|||||: : |:|||| ||::|  ||:::  | |||||  :
Db         64 LIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGNSYSL 123

Qy        121 DNQFHIGDASAVEIKFSNGSQD--------ILLPNVIIMGAEPDLFETNSSNISLR---N 169
              ::|   | ::  : |:   ||          :| |:|| |  | | :       | ||
Db        124 LDKF---DTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDNK 180

Qy        170 NYMPSNHGFGSIAIVTFSPEYSFRFNDN----------SMNEFIQDPALTLMHELIHSLH 219
              || |   |||||  : | |||   | ||             ::: ||||| ||||||| ||
Db        181 NYFPCRDGFGSIMQMAFCPEYVPTF-DNVIENITSLTIGKSKYFQDPALLLMHELIHVLH 239

Qy        220 GLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADY 279
              |||| : :::   |  ||  : :    : ||  |||| | |:|:    ||:|   | ||
Db        240 GLYGMQ-VSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDY 298

Qy        280 KKIASKLSKVQVSNPL---LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK-LYSFT 335
              | ||:|||:|   |     :: || ::: ||  |||::| | || :||  ::   :| ||
Db        299 KAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFT 358

Qy        336 EFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 392
              | :|  || :| | :|    :   |: |||:|:||| :||:||   :||  ::|||  :|
Db        359 EIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVN 418

Qy        393 PRIITPITGRGLVKKIIRFCKNIVSVKGIRKS-------------ICIEINNGELFFVA 438
                : | ||| |:|  || |:    ||::              :||:| | :| |:|
Db        419 TNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFIA 478

Qy        439 SENSYNDDNINTPKEIDDTVTSNN-----NYENDLDQVILNFNSESAPGLSDEKLNLTIQ 493
              :||:::         |: |: |      |:   ||::|:::| :|   | :::     :
Db        479 EKNSFSEEPFQ-----DEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTK 533

Qy        494 NDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYT 553
              | |:| ||  | || |::::  :: || ||| |      : :|:|:| ||:   |||:
Db        534 GIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYS 593

Qy        554 FFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIG 613
              :| |  |: ||:  |  ||: |:: :: ||| |::||:|:|||:|:| :||||| ||||
Db        594 YFPS-VISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIV 652

Qy        614 NEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKE 673
              :  :||| |||  |  :||| ||: :| |   :|     ||  | |:|| |:|  |::
Db        653 KQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAE---SSTQKEKIIKTIDNFLEK 709

Qy        674 RDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNE 733
              | ||| |||  : : |:  :|||| ||  |||:|:|: ||:||| ||: :|  |:  :| :
Db        710 RYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQ 769

Qy        734 LTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLM-KLINEVKINKLREYDENVKTY 792
              : :  :|  ::|:| :| : || ||: |: ||| |:|: ::||| |  :| |:|   |
Db        770 IAD--EINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAK-KQLLEFDTQSKNI 826

Qy        793 LLNYIIQHGSILGESQ-QELNSMVTDTLNNSIPFKLSSYTD---------DKILISYFNK 842
              |: ||  :   :| :: ::| | :    :  |||  |   |          | ||
Db        827 LMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL------ 880

Qy        843 FFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSEVNIS 901
                |::||:   ||   | ||::|::    |   |  |     : |:: ||| :
Db        881 -----KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVH 935

Qy        902 QNDYIIYDNKYKNFSISFWVRIPNYDNKIVNV--NNEYTIINCMRDNN----SGWKVSLN 955
              :   | |:: : ||::|||:|:|      :      |||:||: |: ::    ||| |||
Db        936 KAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLK 995

Qy        956 HNEIIWTLQDNAGINQKLAFNYGNANGISDYI-NKWIFVTITNDRLGDSKLYINGNLIDQ 1014
              | :||||:|:||  ::: |     :  : |: |||:|:|||||||  : ||||| |:
Db        996 GNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGS 1054

Qy       1015 KSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILK 1073
                |  || |   :||  |:  | :  :|: |  | || | |:  ||: ||::  :   |:
Db       1055 AEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLR 1114

Qy       1074 DFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLAN-------------- 1119
              ||||| | || ||||: |   :       ||  : : ||   : | |
Db       1115 DFWGNPLRYDTEYYLIPVASSS------KD--VQLKNITDYMYLTNAPSYTNGKLNIYYR 1166

Qy       1120 RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLFPLYADTATTNKEKTIKI 1177
              |||:|:|  |:|   ::  |: |:  |  ::|:::                | |  :
Db       1167 RLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSY----------------NNNEHIVGY 1210

Qy       1178 SSSGNRFNQVVVMNSVGNNC---------------------TMNFKNNNGNNIGLLGF- 1214
                 || ||  :  :  || |                      :   ::   ::||:|
Db       1211 PKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTH 1270

Qy       1215 --------KADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGW 1249
                     | ::|| ||: |::|      || | |:  : ||
Db       1271 NGQIGNDPNRDILIASNWYFNHLKD--KILGCDWYFVPTDEGW 1311


RESULT 9
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
```

```
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             28.8%;  Score 1883;  DB 1;  Length 1276;
  Best Local Similarity   34.4%;  Pred. No. 4e-69;
  Matches  456;  Conservative  250;  Mismatches  487;  Indels  134;  Gaps   41;

Qy          4 INSFNYNDPVNDRTILYIK---------PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQD 54
              :  |||:|||||  |||::         |       |:| | :|||:||||    | |
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FHPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114
                ||       |||||:|| :||:|| ||| : |:| |||    |  |:  |   :|::
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDNQFHIGDAS---AVEIKFSNGS---QDILLPNVIIMGAEPDLFETNSSNISLR 168
              |: :||:: |      :   ||| || |||    :|: |:|:| |  |:: :  :|
Db        118 GDSSTPEDTFDFTRHTTNIAVE-KFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQG 176

Qy        169 NNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLH 219
                 || |||:::|:  :||:   |:| : |:          |  || : ||||| ||||
Db        177 QQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLH 236

Qy        220 GLYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLA 277
               |||   | :   |  : :    :  | |:  ||  |||| |: ||   : : :     |
Db        237 QLYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALG 295

Qy        278 DYKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLY 332
               || || :|: :  : |       :: || :|  ||  ||| :| : |||:||| ::  |
Db        296 HYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLT 355

Qy        333 S-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNA 389
              :  :|   :::: || |  |  ::      :|:|:|:|| | :|:|:|  |    |||
Db        356 NVMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNI 415

Qy        390 NLNPRIITPITGRGLVKKIIRFCKNIVSVKGIR-KSICIEINNGELFFVASENSYNDDNI 448
               ||  :  ::   :|   : |  :   |  | | ||:: |  | :|| ::| : :
Db        416 ERNP-ALQKLSSESVVDLFTKVCLRL--TKNSRDDSTCIKVKNNRLPYVADKDSISQEIF 472
.
Qy        449 NTPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPK 500
                  |:|   | : :  ||: ||:     | |      | :: | |||  :   :
Db        473 ENKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVF--- 529

Qy        501 YDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFI 560
               ||      || :: |: || ::||::||:      |: ||:|:: ||     |||||  |
Db        530 YD-----DITKY-VDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPS-LA 582

Qy        561 NNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGN 620
               ||| ||| ||::|  :|: ||||   :|  |:|||:|:|:::|||| |||||| | :||
Db        583 EKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGN 642

Qy        621 FKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKE 680
              |  |    |  |||  ||  || : |||   |  |   : |:|| | | |::| ::||:
Db        643 FNQAFATAGVAFLLEGFPEFTIPALGVFT---FYSSIQEREKIIKTIENCLEQRVKRWKD 699

Qy        681 VYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDI 740
               | ::|||:::| ||||     ||| :|  | :|||  |: :|  |:  :|  :  |  :
Db        700 SYQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENI--KSQV 757

Qy        741 KQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQH 800
              : ::| |: |:| |||||::|: | |::|| | :     |::| ::|   || |:| | |
Db        758 ENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSH 817

Qy        801 GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 860
               ||       | : | ::  |::|| : |||:: :|    |::|  |  | :|::: | :
Db        818 NIILVGEVDRLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKN 877

Qy        861 KYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS----EVNISQNDYIIYDNKYKNFS 916
               ||||||:: : :  :|       | |      |||    :: :: |: |:|   |:| |
Db        878 ALVDTSGYNAEVRVGDNVQLNTIYTNDF-----KLSSSGDKIIVNLNNNILYSAIYENSS 932

Qy        917 ISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQK---L 973
              :|||::|      : | :|||||| : :  |||||: : :  | | |||    :|:|   |
Db        933 VSFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQD---VNRKYKSL 985

Qy        974 AFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI 1033
               |:|  :   : | ||| ||||||: :|  ||||||  |   : | :|  : :    |:| |
Db        986 IFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGI 1045

Qy       1034 -VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVL 1092
                |    : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|
Db       1046 DENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN-- 1103
```

```
Qy       1093 KPNNFIDRRKDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSS------TNDN 1140
Db       1104 --DNYIDRYIAP-------ESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDN 1154

Qy       1141 LVRK---NDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVV--VMNSVGN 1195
Db       1155 IILHMLYNSRKYM-IIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSK 1213

Qy       1196 N--CTMNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDHTNSNGCFWNFIS 1244
Db       1214 NKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFIS 1269

Qy       1245 EEHGWQE 1251
Db       1270 RDPGWVE 1276
```

```
RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P.-> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             27.8%;  Score 1820.5;  DB 1;  Length 1290;
  Best Local Similarity   33.7%;  Pred. No. 1.3e-66;
  Matches  449;  Conservative  256;  Mismatches  486;  Indels  141;  Gaps   42;

Qy          4 INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP-- 58
Db          4 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 63

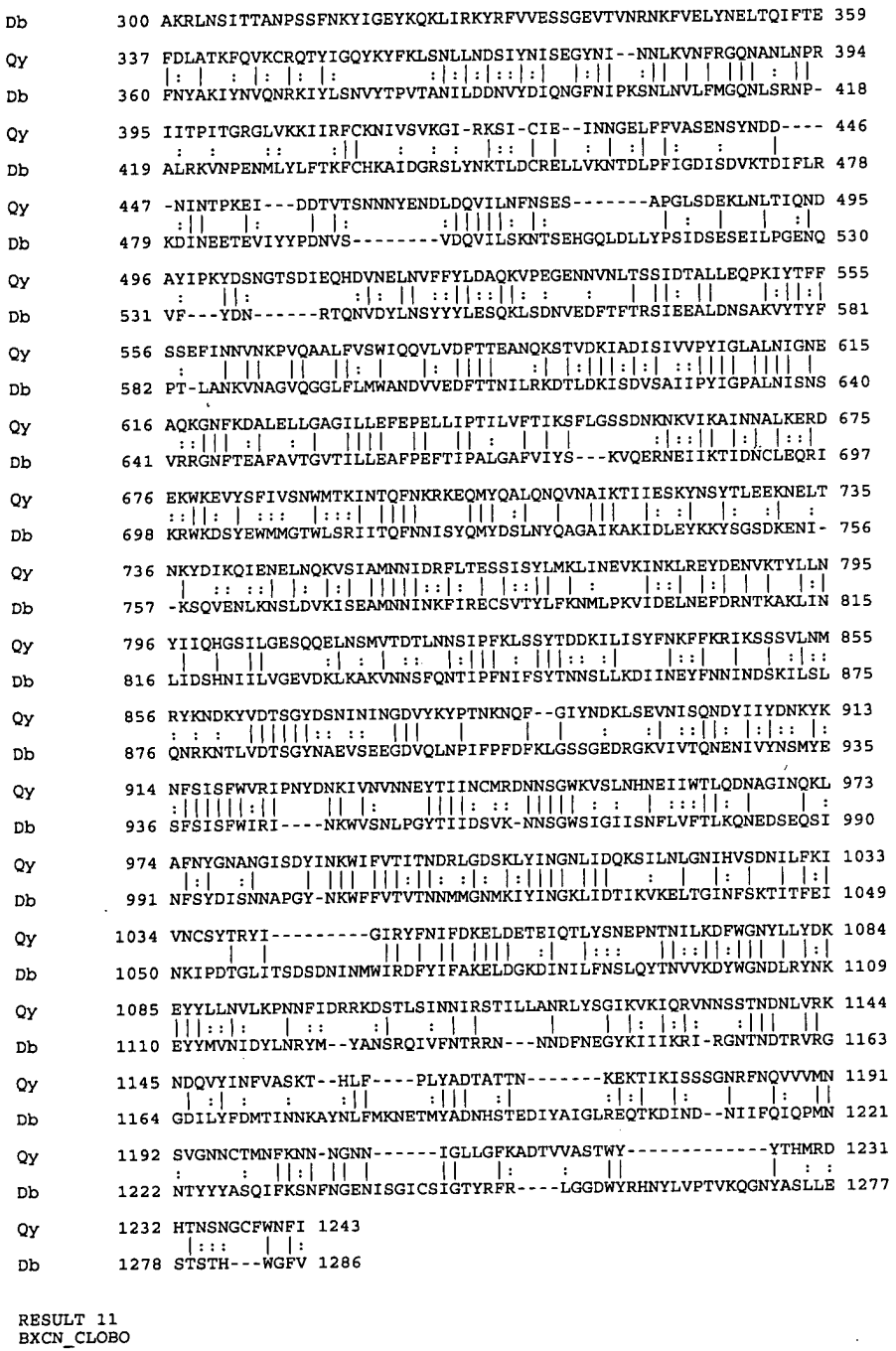
Qy         59 TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDN 118
Db         64 TSPKSG---YYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 120

Qy        119 TPDNQFHIG-DASAVEIKFSNGSQDI----LLPNVIIMGAEPDLFETNSSNISLRNNYMP 173
Db        121 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 180

Qy        174 SNHGFGSIAIVTFSPE----YSFRFND-----NSMNEFIQDPALTLMHELIHSLHGLYGA 224
Db        181 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 240

Qy        225 KGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNLLADYKKI 282
Db        241 -AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 299

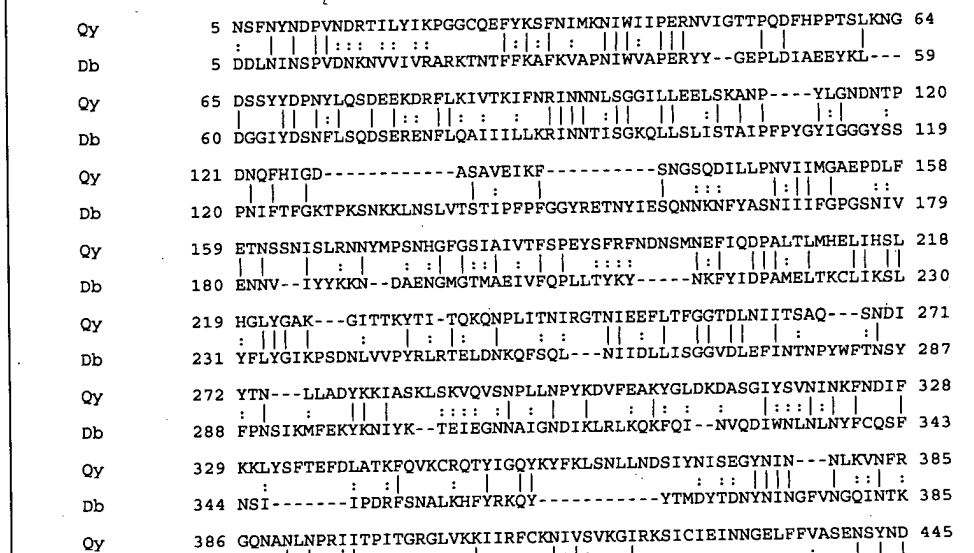
Qy        283 ASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FTE 336
```

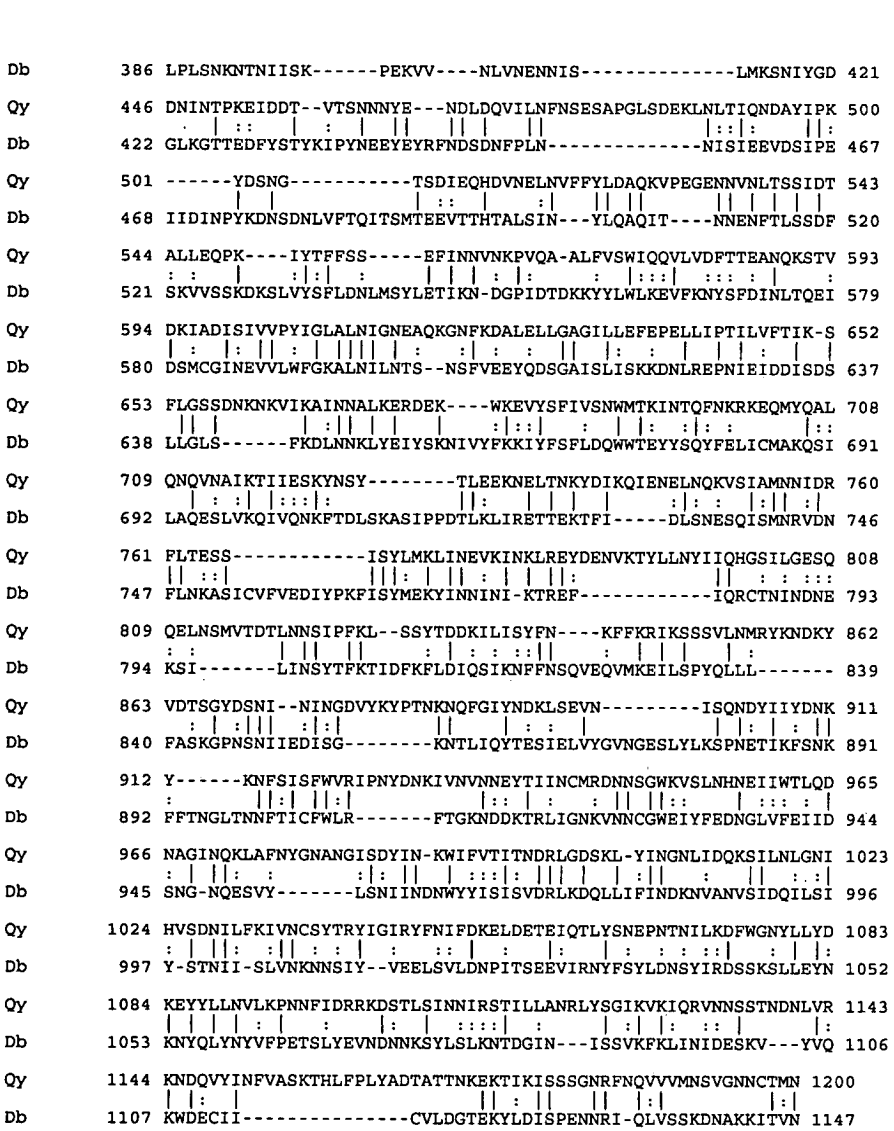


RESULT 11
BXCN_CLOBO

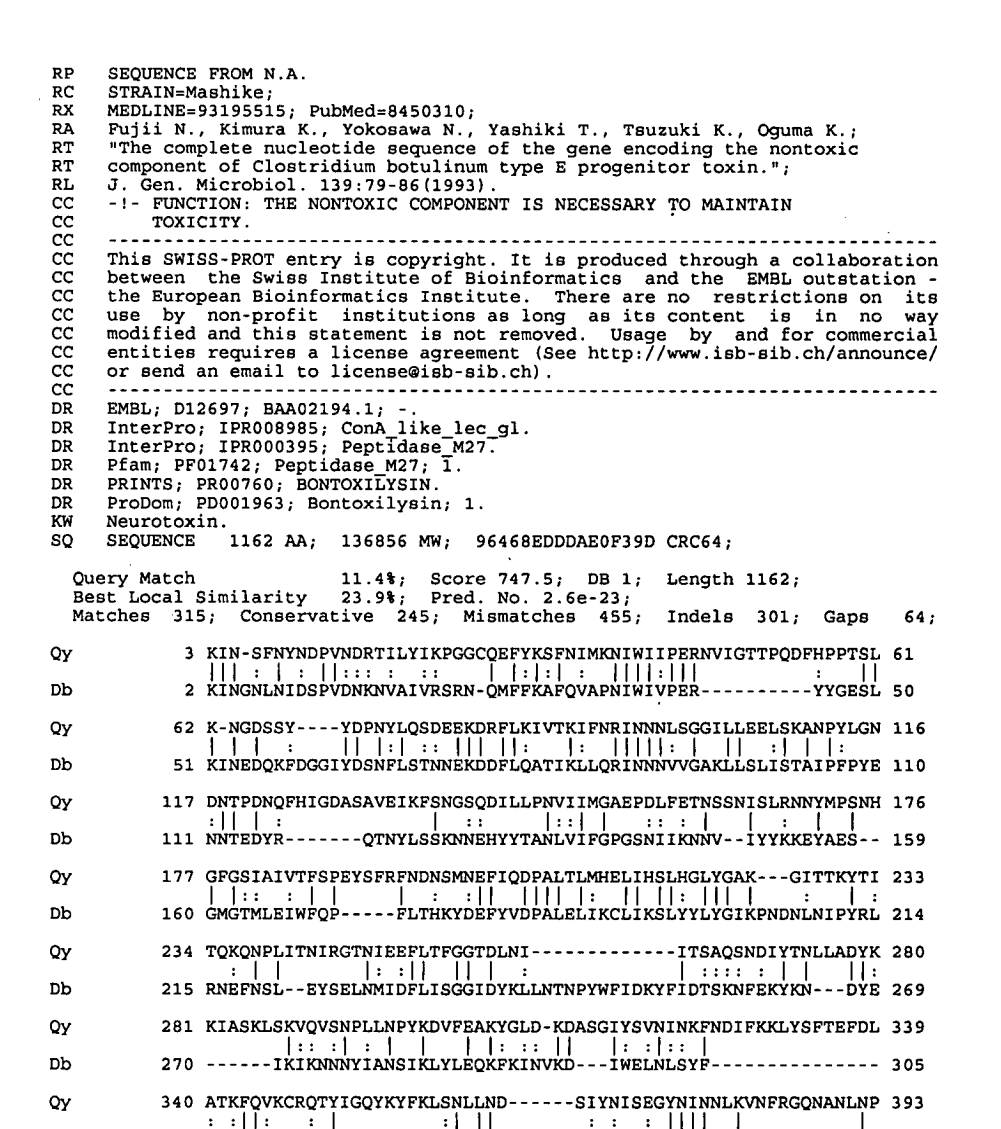
ID BXCN_CLOBO STANDARD; PRT; 1196 AA.
AC P46081;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type C1, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=92231894; PubMed=1567404;
RA Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT "The complete nucleotide sequence of the gene coding for the
RT nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT progenitor toxin.";
RL Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X62389; CAA44262.1; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1196 AA; 138740 MW; 4BD5956274D7F9C3 CRC64;

Query Match 11.6%; Score 761.5; DB 1; Length 1196;
Best Local Similarity 24.1%; Pred. No. 7.4e-24;
Matches 317; Conservative 230; Mismatches 475; Indels 295; Gaps 61;



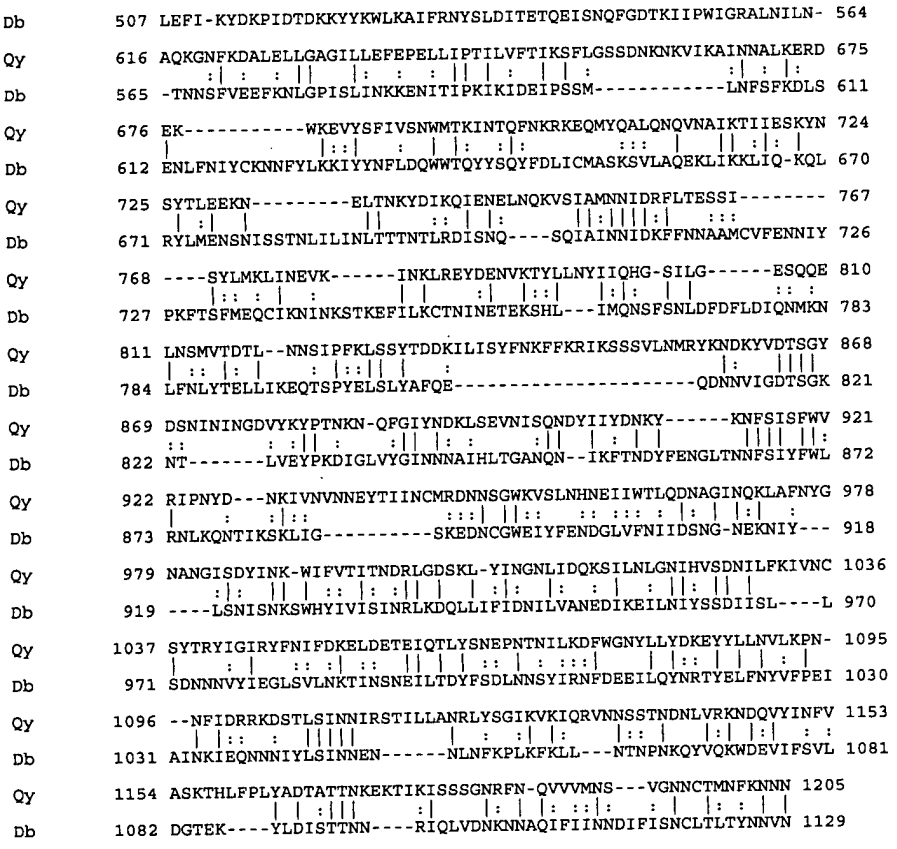


```
RESULT 12
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
```



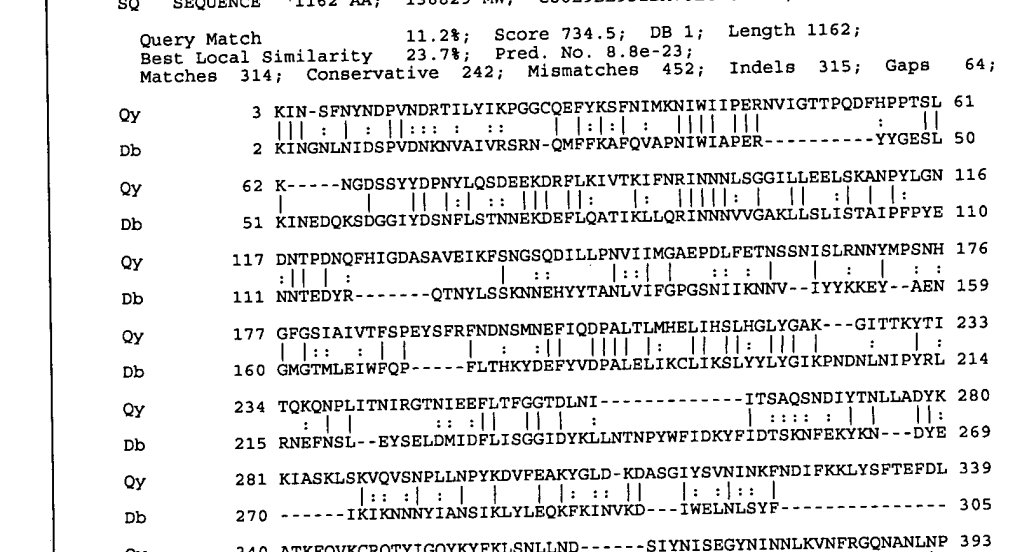
```
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;

  Query Match           11.4%;  Score 747.5;  DB 1;  Length 1162;
  Best Local Similarity  23.9%;  Pred. No. 2.6e-23;
  Matches  315;  Conservative  245;  Mismatches  455;  Indels  301;  Gaps   64;
```



```
RESULT 13
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match           11.2%;  Score 734.5;  DB 1;  Length 1162;
  Best Local Similarity  23.7%;  Pred. No. 8.8e-23;
  Matches  314;  Conservative  242;  Mismatches  452;  Indels  315;  Gaps   64;
```



```
Qy        725 SYTLEEKN---------ELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSI-------- 767
Db        671 RYLMENSNISSTNLILINLTTTNTLRDISNQ----SQIAINNIDKFFNNAAMCVFENNIY 726

Qy        768 ----SYLMKLINEVK------INKLREYDENVKTYLLNYIIQHG-SILG------ESQQE 810
Db        727 PKFTSFMEQCIKNINKSTKEFILKCTNINETEKSHL---IMQNSFSNLDFDFLDIQNMKN 783

Qy        811 LNSMVTDTL--NNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGY 868
Db        784 LFNLYTELLIKEQTSPYELSLYAFQE---------------------QDNNVIGDTSGK 821

Qy        869 DSNININGDVYKYPTNKN-QFGIYNDKLSEVNISQNDYIIYDNKY------KNFSISFWV 921
Db        822 NT-------LVEYPKDIGLVYGINNNAIHLTGANQN--IKFTNDYFENGLTNNFSIYFWL 872

Qy        922 RIPNYD---NKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYG 978
Db        873 RNLNQNTIKSKLIG----------SKEDNCGWEIYFENNGLVFNIIDSNG-NEKNIY--- 918

Qy        979 NANGISDYINK-WIFVTITNDRLGDSKL-YINGNLIDQKSILNLGNIHVSDNILFKIVNC 1036
Db        919 ----LSNISNKSWHYIVISINRLKDQLLIFIDNILVANEDIKEILNIYSSDIISL----L 970

Qy       1037 SYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPN- 1095
Db        971 SDNNNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEI 1030

Qy       1096 --NFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFV 1153
Db       1031 AINKIEQNNNIYLSNNNENSL------NFKPLKFKLL---NTNPNKQYVQKWDEV----- 1076

Qy       1154 ASKTHLFPLYADTATTNKEKTIKISSSGNRFNQV---------VVMNSVG-NNC-TMNFK 1202
Db       1077 -----IFSVLDGT-----EKYLDISIDNNRIQLVDNKNNAKTFIINNDIFISNCLTLTYN 1126

Qy       1203 NNN 1205
Db       1127 NVN 1129
```

```
RESULT 14
YCF2_MARPO
ID   YCF2_MARPO     STANDARD;      PRT;  2136 AA.
AC   P09975;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN   YCF2.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC   Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
CC   -!- SIMILARITY: Belongs to the ycf2 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04465; CAA28078.1; -.
DR   PIR; S01591; A05037.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
KW   Chloroplast; Hypothetical protein.
SQ   SEQUENCE   2136 AA;  259911 MW;  5BD170C9CCF61197 CRC64;

  Query Match             4.5%;  Score 293;  DB 1;  Length 2136;
  Best Local Similarity   20.6%;  Pred. No. 0.00012;
  Matches  303;  Conservative  212;  Mismatches  474;  Indels  480;  Gaps   76;

Qy         91 FNRIN--NNLSGGILLEEL-----SKANPYLGNDNTPDNQFHIGDASAVEIK--FSNGSQ 141
Db        357 FNKFNELNNSEIYIKIEELFSDYIYKFSKYI---------LYEGKKSKTIIKQSFNNNIY 407

Qy        142 DILLPNVIIMGAEPDLFETNSSNIS---LRNNYMPSNHGF-------GSIA---IVTFSP 188
Db        408 YKKLNSIFNFNT---IFYFDSNNLLFDWLKKNYYINNKPFLKSFLIYSSISNQFILFFKQ 464

Qy        189 EYSFRFNDNSM---------NEFIQDPALTLMH---ELIHSLHGLYGAKGITTKYTITQK 236
Db        465 KNSKSFNKNLVKKNSKDVITNVFSKENKIEINNFSKSIYYAFFEILSINEIDNKFVINKI 524

Qy        237 QNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNP-- 294
Db        525 S---LKNINKKKQKRFY------LNKIKSS-DNFRFINLW----KIKNYSSQQFVSNNSF 570

Qy        295 LLNPYKDVFEAKYGLDKDASGIYSVNI---NKFNDIF-----------KKLYSFTEFDLA 340
Db        571 LLNPAFEILQQNYYLKKK-------NILFFKKLNEVFSNFFYFQYYKCKKLNIFLKFASL 623

Qy        341 TKFQVKCRQTYIGQYKYFK--LSNLLNDS-IYNISEGYNINNLKVN-FRGQNANLNPRII 396
Db        624 EKILKKRNKKFTISIKLFKKFYKNKLNENGEYKIESQILQNEKELNKKRKKNFQFNPNI- 682

Qy        397 TPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDD 456
Db        683 ----------KILSF------YNSSKKNIYLQ---NKYFF--NKNLINNKLITWKKISNK 721

Qy        457 TVTSNNNYENDLDQVILN-----FNSESAPGLSD-----------------EKLNLTI 492
Db        722 LVISNSEY----NKIIWNKKNMKFFSFSKNSVLDTFFFNKKSFNIITVIFDKLKKIQLNF 777

Qy        493 QNDAYIPKYDSNGTSDIEQHDVNELNVFF-YLDAQKVPEGEN----NVNLTSSIDTALLE 547
Db        778 QE---------------IQKILNCFSLFFNSKNIKKTKIFKNSYFINENLTT---TFSFN 819

Qy        548 QPKIYTFFSSEFINNVNKPVQAALFVSWI----------------QQVLVDFTTEANQK 590
Db        820 DKEFNIFFLELFISEINNDFLMRFFKKYLYYRIYKDKEILFNPIENRQLLQNFFEKTKIL 879

Qy        591 STVDKIADISIVVPYIGLALNIGN----EAQKGNFK--DALELLGAGILLEFEPELLIPT 644
Db        880 TFIDFLQDPE---------LNYNNRFIFHLEKKTIKNNNLLYLRLLKIFLKDKRNFL--- 927

Qy        645 ILVFTIKSFLGSSDNKNKV-IKA-INNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKE 702
Db        928 -LINEIKSFI---EKKNNLFIKSQLSNVLLV-----KNSYKFFDNIF----NFHFLKQKE 974

Qy        703 QMYQALQNQVNAI-KTIIESKY-------NSYTLEEKNELTNKYDIKQIENELNQKVSIA 754
Db        975 KNIEIILNNQNYFEKSLLKKTYLKNLNLNNSY-----SKFSYKIFIFQLLNILNK----- 1024

Qy        755 MNNIDRFLTESSISYLMKLINEVKINKLRE----YDENV---------------------- 789
Db       1025 -NNYKTFQWISELIFYSKNLNYKIQNKIEKNNYCYNKNISYKKKKIKTVNFFEKNNLFQT 1083

Qy        790 ------------KTYLLNYIIQHGSILGESQQELNSMVTDTL---------NNSIPFKL 827
```

```
Db       1084 NNSWFFTLEWWEYNTYILLQII----------QETFFQITDVLEYFKKKKIIEKNLKFFL 1133

Qy        828 SSYTDDKILISYFN-------KFFKRI--KSSSVLNMRYKNDKYVDTSG---------- 867

Db       1134 KSKKISLKTLSFHNFKLKWNLRFFNEINYKKNYLLNFLWSDFNLINNCNNLYWVIFSLVI 1193

Qy        868 -----YDSNINI-----------NGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNK 911

Db       1194 FIFLYYQKIFSIIIGSDCFHLWKNFEIIQYLTDRSR-SLYFTKLTRRNKTAL------NK 1246

Qy        912 YKNFSISFWVRIPNY------------------DNKIVNVNNE------YTIINCMRDN 946

Db       1247 TENLLSYFFQNLTHYITNIKFYLLTKKNLKKWLINNKTLDLSRRKRKLLVQSLITHNKIQ 1306

Qy        947 NSGWKVSLNHNEIIWT------LQDNAGI--NQKLA--FNYGNANGISDYINKWIFVTIT 996

Db       1307 NYGFE--LNSNKQFFTSYFGYQITNQQGLLYFQYLAQFFQKNLINNSLDLANKWIVFSFW 1364

Qy        997 NDRLGDSKLYINGN---------------LIDQKSILNLGNIHVSDNILFKIVNCSYTRY 1041

Db       1365 HKIFSSQKLRQTNNIELGFQNIPVPLQFGLSYSKGILLIGPIETGRSYLIK--NLAAESY 1422

Qy       1042 IGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL---------------------- 1080

Db       1423 V----------PLFKISINKLLYNKP--DVITESWMNILIESLRRLNLTLDFAKKMSPCI 1470

Qy       1081 LYDKEYYLLNV-----------------LK--PNNFIDRRKDSTLSINN------IRST 1114

Db       1471 IWIQNIHQLNVNRLTQNVESDPTFLLGILLKYFQTDFSKTKKNNIIVIGSTHLPKKVDPA 1530

Qy       1115 ILLANRLYSGIKVKIQRVNNSSTNDNLV--RKNDQV-----YINFVASKTHLFPLYADTA 1167

Db       1531 LISPNRLDKIINVRLFNISQRKKQFPLLLKKKNFQLKENLFFLNEFGSRTMGYNLRDLSA 1590

Qy       1168 TTNKEKTIKISSSGN-----------------------RFNQV--VVMNSVGNNCTMN 1200

Db       1591 LTNEVLLISITKNRSFIDTDTLKLAFHRQIFGLTYTNNKLNFDRIFKIVIYKVGKTIIQN 1650

Qy       1201 --FKNNNGN--NIGLLGFKADTVVASTWY 1225

Db       1651 ILIKSSSMNLLNIGNFLWKKNFYYLSKWY 1679
```

```
RESULT 15
ATX1_PLAFA
ID   ATX1_PLAFA     STANDARD;      PRT;  1956 AA.
AC   Q04956;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cation-transporting ATPase 1 (EC 3.6.3.-).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T9/96;
RX   MEDLINE=93132070; PubMed=8421054;
RA   Krishna S., Cowan G., Meade J.C., Wells R.A., Stringer J.R.,
RA   Robson K.J.;
RT   "A family of cation ATPase-like molecules from Plasmodium
RT   falciparum.";
RL   J. Cell Biol. 120:385-398(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC       ATPases). Subfamily V.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65738; CAA46646.1; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR008250; E1-E2_ATPase_reg.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   TIGRFAMs; TIGR01494; ATPase_P-type; 7.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
FT   DOMAIN       1     35       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    36     58       POTENTIAL.
FT   DOMAIN      59     61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    62     80       POTENTIAL.
FT   DOMAIN      81    407       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   408    427       POTENTIAL.
FT   DOMAIN     428    440       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   441    462       POTENTIAL.
FT   DOMAIN     463   1818       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1819   1837       POTENTIAL.
FT   DOMAIN    1838   1845       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1846   1863       POTENTIAL.
FT   DOMAIN    1864   1881       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1882   1905       POTENTIAL.
FT   DOMAIN    1906   1928       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1929   1952       POTENTIAL.
FT   DOMAIN    1953   1956       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES    496    496       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL     1760   1760       MAGNESIUM (BY SIMILARITY).
FT   METAL     1764   1764       MAGNESIUM (BY SIMILARITY).
FT   DOMAIN     246    251       POLY-ASN.
FT   DOMAIN     252    256       POLY-LYS.
FT   DOMAIN     937    941       POLY-ASN.
FT   DOMAIN    1344   1347       POLY-LYS.
FT   DOMAIN    1363   1372       POLY-ASN.
FT   DOMAIN    1680   1684       POLY-ASN.
SQ   SEQUENCE   1956 AA;  230285 MW;  AE708AAE99009335 CRC64;
```

```
  Query Match             4.2%;  Score 277;  DB 1;  Length 1956;
  Best Local Similarity   19.6%;  Pred. No. 0.00049;
  Matches  309;  Conservative  237;  Mismatches  530;  Indels  498;  Gaps   78;

Qy          3 KINSFNY--NDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Db        256 KINNLNFVKGTYINSNDLLYDDKIGVNIFEDDVNNMKHKF--NQRNI------------ 300

Qy         61 LKNGDSSYY--DPNYLQSDEEK----DRFLKIVTKIFNR---INNNLSGGILLEELSKAN 111

Db        301 -----NYYNKDTNNLEYNNKHRYIYDCLLKKVEAISQKNKIIYSN-------EDINKYM 347

Qy        112 PYLGNDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRN-- 169

Db        348 LYGGT-------YVLSLYNINKIKYNNKEENRILGLVIKTG-----FITTKGKI-VNNIL 394

Qy        170 -------NYMPSNHGFGSIAIV-----TFSPEYSFRFNDNSMN-------EFIQD---PA 207

Db        395 YHKKKELNLINDSYKFLIILIIYALFSVFILLYITLSNNEYTNHIIIKCLDIITDAIPPA 454

Qy        208 L--TLMHELIHSLHGLYGAKGITT----KYTITQKQNPLITNIRGTNIEEFLTFGGTDLN 261

Db        455 LPTTLTVGISIAISRLKKKFSISCLCPHKINIAGQINTMVFDKTGTLTENNLQF----IG 510

Qy        262 IITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVN- 320

Db        511 IITQNKKN---KNMLSDFIHI----------------KEMNTESYIHSKDDNMIHNKNS 550

Qy        321 -INKF--NDIFKKLYSFTEFDLATK----FQVK------CRQTYIGQYK--YFKLSNLLN 365

Db        551 IISEYYIKDNMKNLHTSSKKKSITKERSNFLVQTIKSCLLKDHYIKEKKKEYYTNNTYCN 610

Qy        366 DSIYNIS----------------EGYNINNL-KVNFRGQNANLNPRIITPITGRGLVKKI 408

Db        611 DLHINDSTCSSYLLNSETKDAYCEYYNIDHLCDINKKNMDINSKNELMGKYSKNELMGKT 670
```

```
Qy        409 I--------------RFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKE 453
              |                :: || :  |  :  :  : :  ||    |:|       |
Db        671 IKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKTIKNQVG 730

Qy        454 IDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSD----I 509
              :|  :  : | :||                    |  |    |: |  :|    :
Db        731 VDTNI-YHMNCDND-----------------------YNYDYPCDYNCNNCNDTYHRL 764

Qy        510 EQHDVNELNVFFYLDAQKVP----EGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNK 565
              | |::|: | |       :|    :  ||::    |:  || :        |:  | :
Db        765 EYHNINKDNSF------NIPPEKNKSYNNISEHIKINYPLLFEALACCHTLSKVNNKIMG 818

Qy        566 PVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDAL 625
              |   |  ::      |      |     :|  : |      |    |||   :: :  :
Db        819 DVLEILMFNFTN---CDMLINNNSFIIKEKKKNCSYDFQKIDGDKNIGANDERCHLNN-- 873

Qy        626 ELLGAGILLEFEPELLIPTILVFTIKSFLGSS------------DNKNKVIKAINNALK 672
              |:   ||  || :  :  : |  :||  |::             || :    ||
Db        874 NLVSYNILKRFEFQSRLQRMSVI-VKSTYGNNNDDNNDDDNNNDDDNNDDNNNDDNNDD 932

Qy        673 ERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKY----NSYTL 728
              |:     | : :           | |   :  : |      :|: | : |    | ||
Db        933 NNDDNNNNNYYYNI----------FCKGSPEKIKEL-----CLKSKIPNNYDEILNKYTK 977

Qy        729 EEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLI--NEVKIN------ 780
              :    |:  |  |:::::     :| : |: |   ||:: :|  ||  | :| |
Db        978 QGMRILSISY--KRVKSK-----NINLLNVKRSFVESNLHFLGFLIFTNNMKKNAPDIIH 1030

Qy        781 -------------------------KLREYDENVKTYLLNYIIQHGSILGESQQELNSMV 815
                                       |    : ||:: ::  :|    ::|:: ::   :|
Db       1031 NLQTSGCQCIMSTGDNVLTSIHVAKKCGIINSNVESIIIGDVI---PVVGKNNKQKKKLV 1087

Qy        816 TDTLNNSIPFKLSSYTDDKILI-SYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNI-- 872
              |    |      ||  | : |     :: | ::       |     : |  ||
Db       1088 WYNHKNDTYLK----GHDKTCIDNEFTSIQSQMNSDNIC-----GDNICGDNIYGDNICG 1138

Qy        873 ------NINGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKY-KNFSISFWVRIP- 924
                   |||||       | |   || | ::  ||: ::   |||| |  |::: :    |
Db       1139 DNIYGDNINGD------NINGDNIYGDNINGDNINGDNINTYDNIYGDNYNLDY---CPT 1189

Qy        925 -----NYDNKIVNVNN-----------EYTIINCMRDNNSGWKVSLNHNEI-------IW 961
                  |:| |:  ||             |  |: : :: :          |:|       |
Db       1190 EYHKCTYNNSILYRNNFLYKKENKKDKNYKNISTLYEHRT--------NDIQFDKLCDIL 1241

Qy        962 TLQDNAGINQKL---AF-------------NYGNANGISDYI----NKWIFVTITNDRLG 1001
              :|   :|  |   ||                    |     |  ||    :| |   | |
Db       1242 INKDPRNVNIVLTGKAFIFLKKKFYSFHLPYYEECKNIVHYIMKKKHKKIKNIINNH--- 1298

Qy       1002 DSKLYINGNLIDQKSILNLGNIHVSDN-ILFKI-----VNCSYTRYIGIRYFNIF----- 1050
              :| || : |:||   :  :   ::  | :|:||       |  :  |   :| | :
Db       1299 NSNLYYHYNIID-TFVKRMNKEYMCFNKLLYKIQQKLLYNLIHNLYKKKKYMNNYYDIDE 1357

Qy       1051 ----------------DKELD-ETEIQTLYSNEPNTNIL-----------------K 1073
                            :|:|  : ::: :  || | ||                    |
Db       1358 VHLIGNNNNNNNKNNSKEKKLPLKNKMKHIRKNESNDNITFNTYTSNNIHLSKYKYVHHK 1417

Qy       1074 DFW-----------------GNYLLYDKEYYLLN-VLKPNNF----IDRRKDSTLS-- 1107
              :::                |: |:|: || : :|| :|:    : | ||   |
Db       1418 NYYYPDSCTNLRKKKNSLFYNLKKYIYYEKKKYLQHCLLKHDNYKKVELPRIKDINYSYQ 1477

Qy       1108 ---------INNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVY-------IN 1151
                     |:::      :| : |   :|    :|:  | | :   : :|        |
Db       1478 MESIKTRNFIHSLSEQFAFSNLILSFYIIK--NDDNNVYNKNYIYNKNYIYNKNSICNKN 1535

Qy       1152 FVASKTHLF---PLYADTATTNKE------KTIKISSSGNR----FNQVV---------- 1188
              :: :| :::    :|       ||:      |:: :| |  :    |: ::
Db       1536 YICNKNYIYNKNNIYNKNNIYNKKNILTHAKSVLLSGSSKKFLKFFSNIIRRHKLKEKKN 1595

Qy       1189 -------VMNSVGNN---------CTMNFKNNNG---NNIGLLGFKADTVVASTWYYTHM 1229
                    || | |           ||  || :      |   ::  |   :     |  ||
Db       1596 KKNIKRYKMNHVNNTSKGHIILNMCTHGFKKDYSSLKNKYRIVNNKRYMLKNDNVYDRHM 1655
```

```
Qy       1230 RDHTN-----SNGC 1238
                : |:       ||
Db       1656 YNLTDMYRGTQYGC 1669
```

Search completed: September 10, 2004, 03:35:37
Job time : 25.2761 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 57.2852 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-52
Perfect score: 6541
Sequence: 1 MPKINSFNYNDPVNDRTILY.....TNSNGCFWNFISEEHGWQEK 1252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6361	97.2	1252	2	Q8KZM3	Q8kzm3 clostridium
2	6361	97.2	1255	2	Q9FAR6	Q9far6 clostridium
3	6335.5	96.9	1251	2	Q9K395	Q9k395 clostridium
4	4296.5	65.7	1278	2	Q57236	Q57236 clostridium
5	4204.5	64.3	1268	2	Q45851	Q45851 clostridium
6	4077	62.3	1280	2	Q9ZAJ5	Q9zaj5 clostridium
7	2257	34.5	1291	2	Q8GR96	Q8gr96 clostridium
8	2233	34.1	1291	2	Q933K0	Q933k0 clostridium
9	2229	34.1	1291	2	Q9ZAJ8	Q9zaj8 clostridium
10	2227.5	34.1	1291	2	Q93G71	Q93g71 clostridium
11	2208	33.8	1291	2	Q08077	Q08077 clostridium
12	2143	32.8	1310	2	Q93N27	Q93n27 clostridium
13	1904	29.1	367	2	Q45862	Q45862 clostridium
14	1895	29.0	367	2	Q45861	Q45861 clostridium
15	1880	28.7	1275	12	Q9QTG7	Q9qtg7 clostridium
16	1863	28.5	1285	2	Q45967	Q45967 clostridium
17	1860	28.4	1285	2	Q9LBR1	Q9lbr1 clostridium
18	1822	27.9	1291	2	Q93HT3	Q93ht3 clostridium
19	1787.5	27.3	1280	2	Q9LBS7	Q9lbs7 clostridium
20	1782.5	27.3	1280	2	Q841S3	Q841s3 clostridium
21	1779.5	27.2	1280	2	Q45849	Q45849 clostridium
22	1006.5	15.4	347	2	Q840G4	Q840g4 clostridium
23	880	13.5	173	2	Q9KW90	Q9kw90 clostridium
24	867	13.3	173	2	Q9KW88	Q9kw88 clostridium
25	865	13.2	173	2	Q9KW89	Q9kw89 clostridium
26	858	13.1	173	2	Q9K345	Q9k345 clostridium
27	787	12.0	361	2	Q45848	Q45848 clostridium
28	787	12.0	361	2	Q45846	Q45846 clostridium
29	778	11.9	1163	2	Q45850	Q45850 clostridium
30	766.5	11.7	1196	2	Q53550	Q53550 clostridium
31	766	11.7	1160	2	O87710	O87710 clostridium
32	762.5	11.7	1196	2	Q45916	Q45916 clostridium
33	762.5	11.7	1196	9	Q38197	Q38197 clostridium
34	760.5	11.6	1196	2	Q93HT4	Q93ht4 clostridium
35	756.5	11.6	1196	2	Q9LBS8	Q9lbs8 clostridium
36	756.5	11.6	1196	9	Q9ZX77	Q9zx77 clostridium
37	755	11.5	1193	2	P71108	P71108 clostridium
38	750	11.5	1198	2	O06018	O06018 clostridium
39	747	11.4	1159	2	P71109	P71109 clostridium
40	744.5	11.4	1196	2	Q9LBR2	Q9lbr2 clostridium
41	732.5	11.2	441	2	Q9X708	Q9x708 clostridium
42	724	11.1	1161	2	O69276	O69276 clostridium
43	723.5	11.1	1162	2	Q9ZAJ6	Q9zaj6 clostridium
44	712	10.9	1161	2	Q45891	Q45891 clostridium
45	698	10.7	451	2	Q9R631	Q9r631 clostridium

ALIGNMENTS

RESULT 1
Q8KZM3
ID Q8KZM3 PRELIMINARY; PRT; 1252 AA.
AC Q8KZM3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 5262;
RA Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA Nakamura S., Karasawa T., Kozaki S.;
RT "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT complete cds.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB088207; BAC05434.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1252 AA; 143510 MW; 41B633BB744D3B41 CRC64;

Query Match 97.2%; Score 6361; DB 2; Length 1252;

Best Local Similarity 97.4%; Pred. No. 1.6e-238;
Matches 1219; Conservative 15; Mismatches 18; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||||:||||||||||||||||||| |||| ||||
Db          1 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        121 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        721 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        961 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              |||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              ||||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||
Db       1081 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1141 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1252
```

RESULT 2
Q9FAR6
ID Q9FAR6 PRELIMINARY; PRT; 1255 AA.
AC Q9FAR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX MEDLINE=20509829; PubMed=11055954;
RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT butyricum Strains.";
RL Appl. Environ. Microbiol. 66:4992-4997(2000).
DR EMBL; AB039264; BAB12249.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CD8E4D CRC64;

Query Match 97.2%; Score 6361; DB 2; Length 1255;
Best Local Similarity 97.4%; Pred. No. 1.6e-238;
Matches 1219; Conservative 15; Mismatches 18; Indels 0; Gaps 0;

```
Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              || |||||||||||:|||||||||||||:||||||||||||||||||| |||| ||||
Db          4 MPTINSFNYNDPVNNRTILYIKPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 63

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||||||||||||||:|||:|||||||||||||:|||| ||||||||||||||||||||
Db         64 LKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNTP 123

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              |  | | ||||| |:|||||| ||||||||||||||||||||||||||||||||||||
Db        124 DGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 183

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        184 IAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 243

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 303

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 363

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 423

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 483

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 543

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        544 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADIS 603

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 663

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:| |||
Db        664 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIE 723

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db        724 SKYNSYTLEEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 783

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||:|||||||||||||||| ||||||||||||||||||||||||||
Db        784 KLREYDENVKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 843

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 903

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 963

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||:||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        964 WTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNL 1023

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db       1024 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYL 1083

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              |||||||||||||||||:|| ||||||||||||||||||||||||||||||||||||||||
Db       1084 LYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1143

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1144 LVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1203

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              |||||||||||||||||||||||||||||:||||| ||||||||||||||
Db       1204 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISEEHGWQEK 1255


RESULT 3
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
```

```
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             96.9%;  Score 6335.5;  DB 2;  Length 1251;
  Best Local Similarity   97.0%;  Pred. No. 1.6e-237;
  Matches 1214;  Conservative   15;  Mismatches   22;  Indels    1;  Gaps    1;

Qy          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120

Qy        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS 180

Qy        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL 240

Qy        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK 300

Qy        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL 360

Qy        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG 420

Qy        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA 480

Qy        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS 540

Qy        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS 600

Qy        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK 660
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        601 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFVPELLIPTILVFTIKSFLGSSDNK 660

Qy        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        661 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKTIIE 720

Qy        721 SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780
              |||||||||| || | |||:|||||||||||||||||||||||||||||||||||||||||
Db        721 FKYNSYTLEEKKELKNNYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN 780

Qy        781 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF 840
              |||||||||||||:|||||||||||||||||| ||||||||||||||||||||||||||||
Db        781 KLREYDENVKTYLLDYIIQHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYF 840

Qy        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI 900
              |||||||||||||||||||||||||||||||||||||||:::|||||||| |:| | |||||
Db        841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFNSKPSEVNI 900

Qy        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII 960
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        901 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNINNEYTIINCMRDNNSGWKVSLNHNEII 960

Qy        961 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL 1020
              ||||||| ||||| | ||||||||||||||||||||||||||||||:|||||||||||
Db        961 WTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGHLIDQKSILNL 1020

Qy       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140
              |||| |||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db       1081 LYDKGYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNDSSTNDR 1140

Qy       1141 LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN 1200
              ||||||||:: | :  : ||||| ||:|||||| ||||||||||||||||||||||||||
Db       1141 FVRKNDQVYINYI-SNSSSYSLYADTNTTDKEKTIKSSSSGNRFNQVVVMNSVGNNCTMN 1199

Qy       1201 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK 1251


RESULT 4
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           65.7%;  Score 4296.5;  DB 2;  Length 1278;
  Best Local Similarity  64.0%;  Pred. No. 1.3e-158;
  Matches  818;  Conservative  207;  Mismatches  218;  Indels   35;  Gaps   14;

Qy          4 INSFNYNDPVNDRTILYIK---PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              |||||||||||| ||||::      :::||:| ||:|:|||||| ||| | || || |
Db          5 INSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFDPPAS 64

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |:|| |:|||||| :| ||||:||   |:| |||:| :| :||:|:| | ||||||::||
Db         65 LENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGNEHTP 124

Qy        121 DNQFH-IGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNI-SLRNN---YMPSN 175
               |:|| :   ::| || |   :  :: |::::|| ||:|| :| :  | ::   | |||
Db        125 INEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVYDPSN 184

Qy        176 HGFGSIAIVTFSPEYSFRFND------NSMNEFIQDPALTLMHELIHSLHGLYGAKGITT 229
              |||||  |||||||| : |||      :|   || |||::| |||||:|||||||:|:|
Db        185 DGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGARGVTY 244

Qy        230 KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKV 289
              | ||  || ||:   :   :|||||||| ||||||||    || ||||:|:|||::||:|
Db        245 KETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATRLSRV 304

Qy        290 QVSNPL--LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKC 347
               : |   :| ||| |: |||||:| | |:|| |||||:|:|||||||| ||| ||:|||
Db        305 NSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKFKVKC 364

Qy        348 RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKK 407
              | ||  :| : |: |||:| || :|||:|| || || |||| |||:|| |  :|||:|
Db        365 RNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKGLVEK 424

Qy        408 IIRFCKNIVSVKGIR--KSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYE 465
              |::|||:::  || :   :|| :|| |||||||:|||:::|||||||||   ||||
Db        425 IVKFCKSVIPRKGTKAPPRLCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLNNNYR 484

Qy        466 NDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDA 525
              |:||:|||::|||: | :|:: ||  :|:|:|:|:|||||||:||:|:| :|||||| |
Db        485 NNLDEVILDYNSETIPQISNQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFFYLHA 544

Qy        526 QKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTT 585
              ||||||| |::|||||||| |: ::|||||||||| :|||| ||||:||| ||: ||||
Db        545 QKVPEGETNISLTSSIDTALSEESQVYTFFSSEFINTINKPVHAALFISWINQVIRDFTT 604

Qy        586 EANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTI 645
              || |||| |||||||:||||:||||||||| || |||:| |||||||||| ||||||||
Db        605 EATQKSTFDKIADISLVVPYVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELLIPTI 664

Qy        646 LVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMY 705
              |||||||:|||:||||:||||||:| ||: ||||:||:|||||:|:|||||||||||||
Db        665 LVFTIKSFIGSSENKNKIIKAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMY 724

Qy        706 QALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTES 765
              |||||||:||||:|| |||:|| :|:| | ::|:|   |  |||:|||:|| ||:||:|||
Db        725 QALQNQVDAIKTVIEYKYNNYTSDERNRLESEYINNIREELNKKVSLAMENIERFITES 784

Qy        766 SISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPF 825
              || |||||||| |::|||||| || |||:|| :| |||| | |||| :|| ||||||||
Db        785 SIFYLMKLINEAKVSKLREYDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNNSIPF 844

Qy        826 KLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNK 885
              :|||||:||||| |||| :|:|| :|:|:|||:|:|:::| ||| |||:|||||| | ||:
Db        845 ELSSYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNR 904

Qy        886 NQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRD 945
              ||||||: | |||||:||: ||||: :|:|||||||||| | || ||:|||||||:|:|:
Db        905 NQFGIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRN 963

Qy        946 NNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKL 1005
              ||||||:|||:|:||||||| || |||| |||     |||||||||||||||:|||:|::
Db        964 NNSGWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRI 1023

Qy       1006 YINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSN 1065
              |||||||:||| |||:|||||||||||||| |: |||:||||| :|| || :|||:||||:
Db       1024 YINGNLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSD 1083

Qy       1066 EPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRS-----TILLANR 1120
              ||: :|||||||||||:| |||||:|: :  |  :  : |:||  |       |    |
Db       1084 EPDPSILKDFWGNYLLYNKRYYLLNLLRTDKSI-TQNSNFLNINQQRGVYQKPNIFSNTR 1142

Qy       1121 LYSGIKVKIQRVNNSSTN----DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIK 1176
              ||:|::| |::  | ||:    || |||||  ||| |  :   : |||| :    || ||
Db       1143 LYTGVEVIIRK--NGSTDISNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIK 1199

Qy       1177 I---SSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHT 1233
              :   |:| |   |::||:|||||||:|||| |||||| .: :|||:||| ::| :|
Db       1200 LIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNT 1259

Qy       1234 NSNGCFWNFISEEHGWQE 1251
              :||||||:|||:||||||
Db       1260 SSNGCFWSFISKEHGWQE 1277


RESULT 5
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match           64.3%;  Score 4204.5;  DB 2;  Length 1268;
  Best Local Similarity   63.1%;  Pred. No. 4.8e-155;
  Matches  803;  Conservative  219;  Mismatches  225;  Indels   25;  Gaps   14;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:||||||:|: ||||:|       ::||:| || |:||||||:||  | ||:
Db          1 MPVNINNFNYNDPINNTTILYMKMPYYEDSNKYYKAFEIMDNVWIIPERNIIGKKPSDFY 60

Qy         57 PPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGN 116
              || || :| |:|||||||  :| ||||||| | |:|||||:| :| :||||:     |||||
Db         61 PPISLDSGSSAYYDPNYLTTDAEKDRFLKTVIKLFNRINSNPAGQVLLEEIKNGKPYLGN 120

Qy        117 DNTPDNQFHIGDAS-AVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNN---YM 172
              |:|  |:|   : | :|||| |||: | :| |::|:|  |:: | ::  : :  |   |
Db        121 DHTAVNEFCANNRSTSVEIKESNGTTDSMLLNLVILGPGPNILECSTFPVRIFPNNIAYD 180

Qy        173 PSNHGFGSIAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYT 232
              ||  ||||| :::|| || : |||| : || |||::| ||||| |||||||||:| |
Db        181 PSEKGFGSIQLMSFSTEYEYAFNDNT-DLFIADPAISLAHELIHVLHGLYGAKGVTNKKV 239

Qy        233 ITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVS 292
              |   |  |:   :   ||||:|||| ||||||:: :  ||  ||::|  |||:||:|  :
Db        240 IEVDQGALMAAEKDIKIEEFITFGGQDLNIITNSTNQKIYVILLSNYTAIASRLSQVNRN 299

Qy        293 NPLLNP--YKDVPEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQT 350
              |  ||   ||: |: |||||:|::| |:|||:||| |:|||:|||| ||| ||||| |
Db        300 NSALNTTYYKNFFQWKYGLDQDSNGNYTVNISKFNAIYKKLFSFTECDLAQKFQVKNRSN 359

Qy        351 YIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIR 410
              |:  :| |:| :||:|:||:||||:|| :|:||  ||| ||| ||: ||   |||:: :
Db        360 YLFHFKPFRLLDLLDDNIYSISEGFNIGSLRVNNNGQNINLNSRIVGPIPDNGLVERFVG 419

Qy        411 FCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQ 470
              ||:||| || : |:||::|| :|||||||:|||:: ||:|||||||  :||||: :||:
Db        420 LCKSIVSKKGTKNSLCIKVNNRDLFFVASESSYNENGINSPKEIDDTTITNNNYKKNLDE 479

Qy        471 VILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPE 530
              |||::||:: | ||   || | |||:|:|||||||||:|::: |::|||||| ||| ||
Db        480 VILDYNSDAIPNLSSRLLNTTAQNDSYVPKYDSNGTSEIKEYTVDKLNVFFYLYAQKAPE 539

Qy        531 GENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQK 590
              ||: ::||||::|||:  |:|||||:||| ||||||||||:|||||: |||||| ||
Db        540 GESAISLTSSVNTALLDASKVYTFFSSDFINTVNKPVQAALFISWIQQVINDFTTEATQK 599

Qy        591 STVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTI 650
              ||:|||||||::|||:|||||||| ||||||:|:|||||||||| ||||||||||||
Db        600 STIDKIADISLIVPYVGLALNIGNEVQKGNFKEAIELLGAGILLEFVPELLIPTILVFTI 659

Qy        651 KSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQN 710
              |||: | |:|||:|||||||||:|||: |||||||:|||||:|:|||||||||||||||
Db        660 KSFINSDDSKNKIIKAINNALRERELKWKEVYSWIVSNWLTRINTQFNKRKEQMYQALQN 719

Qy        711 QVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYL 770
              ||: || ||| |||:|||:||| |  :|:|  |: |||:|||:|| ||||||||||||||
Db        720 QVDGIKKIIEYKYNNYTLDEKNRLRAEYNIYSIKEELNKKVSLAMQNIDRFLTESSISYL 779

Qy        771 MKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQ-QELNSMVTDTLNNSIPFKLSS 829
              |||||| ||||| |||: |  |||||||:: | || |    |||:|::|||||||:||
Db        780 MKLINEAKINKLSEYDKRVNQYLLNYILENSSTLGTSSVPELNNLVSNTLNNSIPFELSE 839

Qy        830 YTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFG 889
              ||:|||||    :|:||| ||:|||:|:|::::|:||| |||:||||:| | ||:||||
Db        840 YTNDKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFG 899

Qy        890 IYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSG 949
              ||: :||||||:||: |||:::|:|||:|||||| |:| : |:|||||||||||:||||
Db        900 IYSSRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN-LKNLNNEYTIINCMRNNNSG 958

Qy        950 WKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYING 1009
              ||:|||:| |||||||  | |||| |||      |||||||| ||||||:||| ||||||
Db        959 WKISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYING 1018

Qy       1010 NLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069
              || |||||||||||| |||||||| |: |||:|||| ||: |||:|||:||| :||::
Db       1019 NLTDQKSILNLGNIHVDDNILFKIVGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIR----STILLAN-RLYSG 1124
               |||||||||||:|:||||:||||  : :  |  |:||  |     |  : :| |||:|
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPNMSVTKNSD-ILNINRQRGIYSKTNIFSNARLYTG 1137

Qy       1125 IKVKIQRVNNSSTN--DNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIK---ISS 1179
              ::| |::| :: |:  || ||||| |||| |   :  : |||| :|:  |||||    ||:
Db       1138 VEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSE-YQLYADVSTSAVEKTIKLRRISN 1196

Qy       1180 SGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCF 1239
              |    ||:::|:|:|:|||||| ||||:||||||  : :|||:||| ::|::| :||||
Db       1197 SNYNSNQMIIMDSIGDNCTMNFKTNNGNDIGLLGFHLNNLVASSWYYKNIRNNTRNNGCF 1256

Qy       1240 WNFISEEHGWQE 1251
              |:|||:||||||
Db       1257 WSFISKEHGWQE 1268


RESULT 6
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
```

```
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             62.3%;  Score 4077;  DB 2;  Length 1280;
  Best Local Similarity   61.0%;  Pred. No. 4.2e-150;
  Matches  781;  Conservative  223;  Mismatches  240;  Indels   36;  Gaps   15;

Qy          4 INSFNYNDPVNDRTILYI-KP--GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              |||||||||||| |||| ||    :::||:| || |:||:|||: ||| | :|  | |
Db          5 INSFNYNDPVNDETILYMQKPYEERSRKYYKAFEIMPNVWIMPERDTIGTKPDEFQVPDS 64

Qy         61 LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP 120
              |||| |:|||||||| :| |||:|| : |:|||||:| :| :||||:| | ||||:|:|
Db         65 LKNGSSAYYDPNYLTTDAEKDRYLKTMIKLFNRINSNPTGKVLLEEVSNARPYLGDDDTL 124

Qy        121 DNQF-HIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNN----YMPSN 175
               |:|  :   ::| ||||   :  :: |::::|| ||:|:   : :   |       |||
Db        125 INEFLPVNVTTSVNIKFSTDVESSIISNLLVLGAGPDIFKAYCTPLVRFNKSDKLIEPSN 184

Qy        176 HGFGSIAIVTFSPEYSFRFND------NSMNEFIQDPALTLMHELIHSLHGLYGAKGITT 229
              |||||| |:|||||   |||      ||   || |||::| |||||:|||||||| :|
Db        185 HGFGSINILTFSPEYEHIFNDISGGNHNSTESFIADPAISLAHELIHALHGLYGAKAVTH 244

Qy        230 KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKV 289
              | ::  :: ||:   :   :|||||||| ||||| ||    || :|||:|:|||::| :|
Db        245 KESLVAERGPLMIAEKPIRLEEFLTFGGEDLNIIPSAMKEKIYNDLLANYEKIATRLREV 304

Qy        290 QVSNP--LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKC 347
               : |   :| ||| |: |||||::| | |:|| ||||:|:|||||||| ||| ||:|||
Db        305 NTAPPGYDINEYKDYFQWKYGLDRNADGSYTVNRNKFNEIYKKLYSFTEIDLANKFKVKC 364

Qy        348 RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKK 407
              | ||  :| : |: |||:| || :|||:|| || || |||| ||||:||  |  :|||:|
Db        365 RNTYFIKYGFVKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNINLNPKIIDSIPDKGLVEK 424

Qy        408 IIRFCKNIVSVKGIRK--SICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYE 465
              ||:|||:|:  || ::  |:|| :|| |||||||:||| :|||||||||   ||||
Db        425 IIKFCKSIIPRKGTKQSPSLCIRVNNRELFFVASESSYNESDINTPKEIDDTTNLNNNYR 484

Qy        466 NDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDA 525
              |:||:|||::|||: | :|:  ||  :|:::|:|:|||||||::||:: | ||||| |
Db        485 NNLDEVILDYNSETIPQISNRTLNTLVQDNSYVPRYDSNGTSEIEEYDVVDFNVFFYLHA 544

Qy        526 QKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTT 585
              ||||||| |::|||||||||||: |:|||||||||: :|||| ||||: || :|: ||||
Db        545 QKVPEGETNISLTSSIDTALLEESKVYTFFSSEFIDTINKPVNAALFIDWISKVIRDFTT 604

Qy        586 EANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTI 645
              || |||||||||||::|||:|||||  ||:|||||:| ||||||||||| ||| || |
Db        605 EATQKSTVDKIADISLIVPYVGLALNIVIEAEKGNFEEAFELLGAGILLEFVPELTIPVI 664

Qy        646 LVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMY 705
              |||||||:: | :|||| ||||||:| ||: ||||:||:|||||:|:|||||||||||
Db        665 LVFTIKSYIDSYENKNKAIKAINNSLIEREAKWKEIYSWIVSNWLTRINTQFNKRKEQMY 724

Qy        706 QALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTES 765
              |||||||:|||| || |||:|| :||| | :||:|  || |||:|||:|| ||:||:|||
Db        725 QALQNQVDAIKTAIEYKYNNYTSDEKNRLESKYNINNIEEELNKKVSLAMKNIERFMTES 784

Qy        766 SISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPF 825
              |||||||||| :: ||:|||::||: ||:||: |  |||| :||  :|| |||:||||
Db        785 SISYLMKLINEAEVGKLKEYDKHVKSDLLDYILYHKLILGEQTKELIDLVTSTLNSSIPF 844

Qy        826 KLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNK 885
              :|||||:||||| ||| : :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:
Db        845 ELSSYTNDKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNR 904

Qy        886 NQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRD 945
              ||||||: :||||||:||: |||:::|:|||||||| || :  : :| | |||||||| :
Db        905 NQFGIYSGRLSEVNIAQNNDIIYNSRYQNFSISFWVTIPKH-YRPMNRNREYTIINCMGN 963

Qy        946 NNSGWKVSL---NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGD 1002
              ||||||:||      |||||||| :|  :|| | |     |||||||||||||||:|||:
Db        964 NNSGWKISLRTIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGN 1023

Qy       1003 SKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTL 1062
              |::||||||| :||| |||:|||||||||||| |    |:|||| :|: |||:|||:||
Db       1024 SRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETL 1083

Qy       1063 YSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRS-----TILL 1117
              |||||: :||||:|||||||:|:||| |:|: : :|  |    |:|| |       :: |
Db       1084 YSNEPDPSILKDYWGNYLLYNKKYYLFNLLRKDKYI-TRNSGILNINQQRGVTGGISVFL 1142

Qy       1118 ANRLYSGIKVKIQR---VNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKT 1174
                :|| |::| |::   :: |:| || |||||  ||| |      : |||| : |  ||
Db       1143 NYKLYEGVEVIIRKNAPIDISNT-DNFVRKNDLAYINVVDHGVE-YRLYADISITKSEKI 1200

Qy       1175 IKI---SSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRD 1231
              ||:   |:  :   |::||:|:|||||||:||:|:|||||| :| :|||:||| |:|
Db       1201 IKLIRTSNPNDSLGQIIVMDSIGNNCTMNFQNNDGSNIGLLGFHSDDLVASSWYYNHIRR 1260

Qy       1232 HTNSNGCFWNFISEEHGWQE 1251
              :|:||||||:|||:||||:|
Db       1261 NTSSNGCFWSFISKEHGWKE 1280


RESULT 7
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
```

```
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             34.5%;  Score 2257;  DB 2;  Length 1291;
  Best Local Similarity   38.2%;  Pred. No. 1.2e-79;
  Matches  502;  Conservative  269;  Mismatches  458;  Indels   84;  Gaps   33;

Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||:::  |: ::|     |   :||:| |   |||||||   | |:||:
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
                : : | |   ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
              :   | :|:   ||   |  :   ::     :  |:|| |  | | |  : :| ::|:
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
              :  | ||| |  : | |||   ||:   |:         :  |||| ||||||| ||||
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKK 281
              || | :     :  ::   : :       ||  |||| | :||| :    ||  :| :::
Db        240 YGIK-VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRG 298

Qy        282 IASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTE 336
              |  :|:|| |      |  :| ||: |: ||    :|: | ||:::  |: ::| | : |||
Db        299 IVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTE 358

Qy        337 FDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNP 393
              ::|  :::| | :|         |: |||:: || | ||:||:  |::  :||||  :|
Db        359 TNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINK 418

Qy        394 RIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDD-NINTPK 452
              :    |:   |    |: ||:: :       |||:::| :|||:| :|||::|| : |
Db        419 QAYEEISKEHLAVYKIQMCKSVRA-----PGICIDVDNEDLFFIADKNSFSDDLSKNERI 473

Qy        453 EIDDTVTSNNNYEN--DLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIE 510
              | |   | :|  ||   :|::||: |   |   |  |         :  :| |:       |:
Db        474 EYD---TQSNYIENRSSIDELILDTNLISKIELPSENTESLTDFNVDVPVYEKQPA--IK 528

Qy        511 QHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAA 570
              :   :|  :| || :|  |    :::|||| | |||    |:|:||| ::|   || |:|
Db        529 KIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAG 588

Qy        571 LFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGA 630
              ||  |::|:: ||  |||: ||:|||||||::||||||||:|||  ||||::| |: ||
Db        589 LFAGWVKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGA 648

Qy        631 GILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWM 690
              ||||| |||||  :  | ::|::   |||||:|| |:||| :|||| ::|  ||: |:
Db        649 SILLEFIPELLIPVVGAFLLESYI---DNKNKIIKTIDNALTKRDEKWIDMYGLIVAQWL 705

Qy        691 TKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQK 750
              : :||||   || ||:||  |   |:: ||: ||| |: :||: :    |   | ::||:
Db        706 STVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNIYSEKEKSNI--NIDFNDINSKLNEG 763

Qy        751 VSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQE 810
              :: |::||: |: | |:|||||  :  : : || ::|  :|  |||||  ::    |  | :
Db        764 INQAVDNINNFINECSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEY 823

Qy        811 LNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDS 870
              | |   |  ||| || ||:: |||  |||:   | :: :||:||:::  :| ||| :
Db        824 EKSKVDKHLKTIIPFDLSMYTNNTILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGA 883

Qy        871 NININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKI 930
              |: :   |     :|||| : :   ||: ::||  ||::: : :||:|||:||| | |
Db        884 NVEVYDGVEL--NDKNQFKLTSSTNSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDG 941

Qy        931 VN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYIN 988
              :   ::|||||||:: |||||:|:  | ||||| |   |  : : | |      ||||||
Db        942 IQNYIHNEYTIINCIK-NNSGWKISIRGNRIIWTLTDINGKTKSVFFEYSIREDISDYIN 1000

Qy        989 KWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYF 1047
              :| ||||||:   ::|:|||| |     | ::| :  :  |:||:  :   |::| ::||
Db       1001 RWFFVTITNNS-DNAKIYINGKLESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYF 1059

Qy       1048 NIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS--- 1104
              :||: || :: |: :|  :  :  |||||||  |:|:|||||: |    |::|  :|||
Db       1060 SIFNTELSQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVG 1119

Qy       1105 ---TLSINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFVASKTHLF 1160
                 | |  |  |  :    || | |  |:| :|| | ||::||| | :|::|  |    :
Db       1120 EILTRSKYNQNSNYINYRNLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNSNRE-W 1178

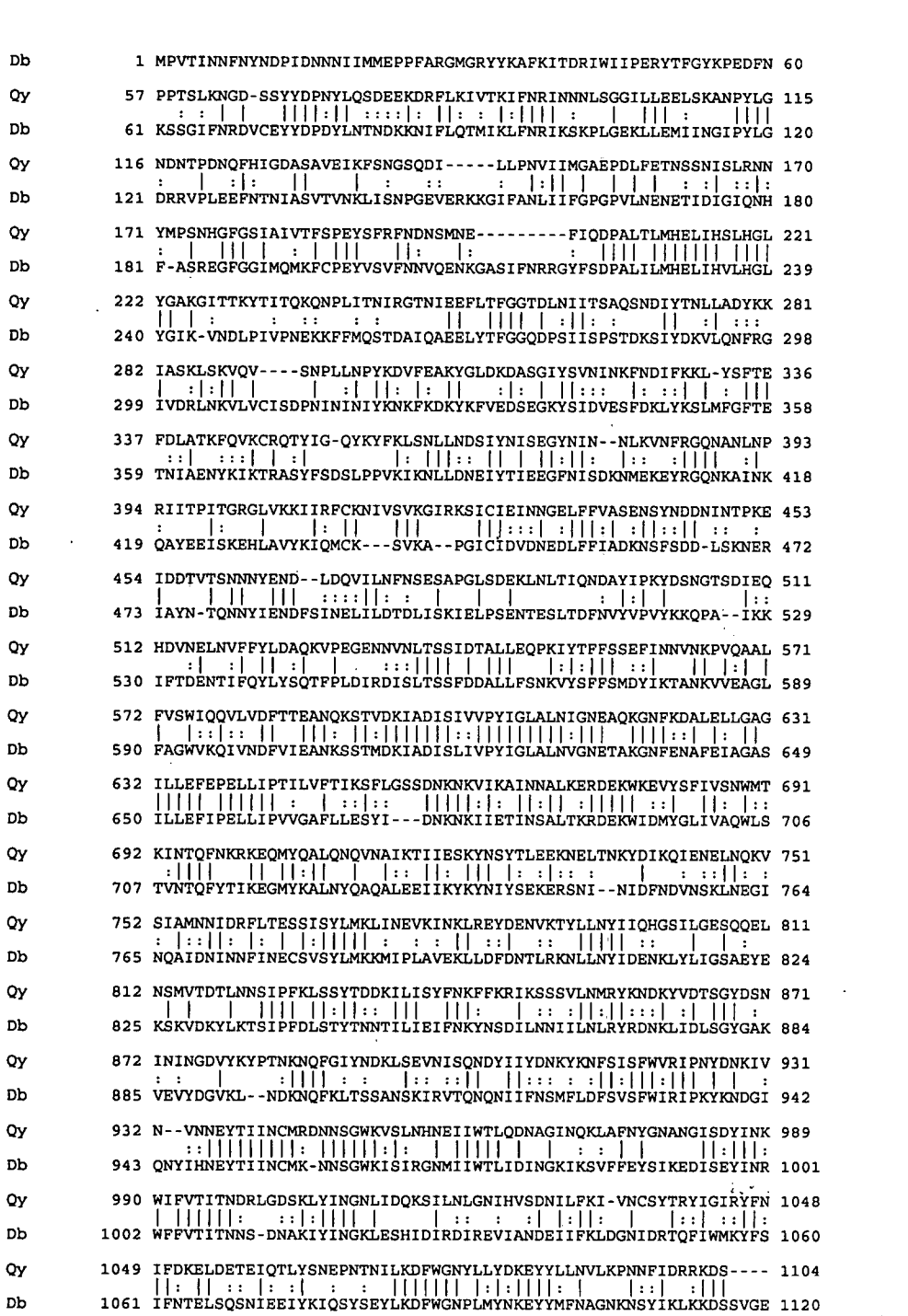
Qy       1161 PLYADTATTNKEKTIKISS--SGNRFNQVVVMNSVGN----NCTMNFKNN--NGNNIGLL 1212
              :||     :|| : :::    | | : : :          :| : || :  : : |||:
Db       1179 RVYAYKDFKEEEKKLFLANIYDSNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLI 1238

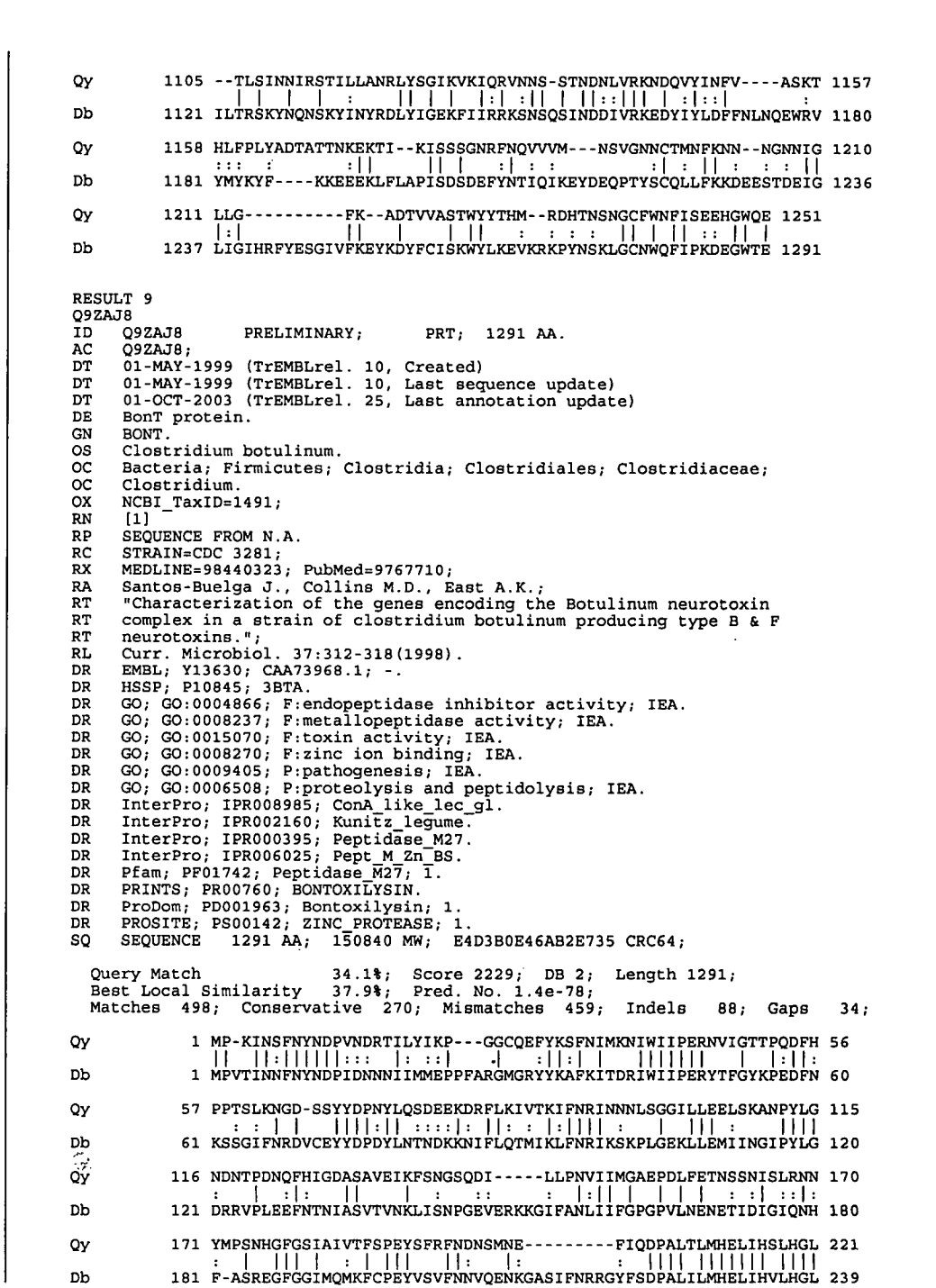
Qy       1213 G----FKADTVV--------ASTWYYTHM-RDHTNSN-GCFWNFISEEHGWQE 1251
              |    :::  |:          | ||   : |  | | || | || :: || |
Db       1239 GIHRFYESGIVLKDYKNYFCISKWYLKEVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 8
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             34.1%;  Score 2233;  DB 2;  Length 1291;
  Best Local Similarity   37.9%;  Pred. No. 1e-78;
  Matches  498;  Conservative  272;  Mismatches  457;  Indels   88;  Gaps   34;

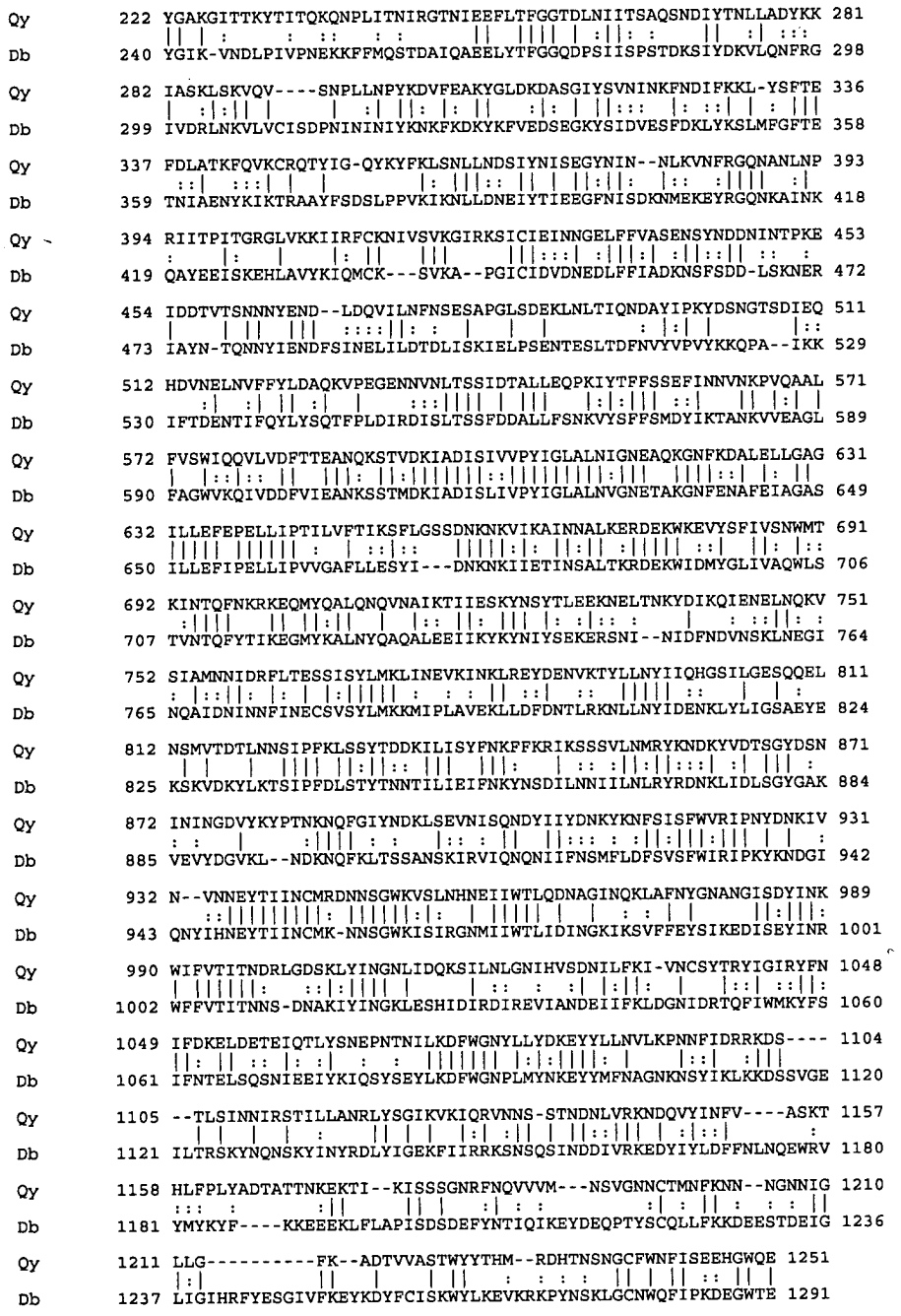
Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
              ||  ||:|||||:::  |: ::|     |   :||:| |   |||||||   | |:||:
```





RESULT 9
Q9ZAJ8
ID Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 34.1%; Score 2229; DB 2; Length 1291;
Best Local Similarity 37.9%; Pred. No. 1.4e-78;
Matches 498; Conservative 270; Mismatches 459; Indels 88; Gaps 34;



```
RESULT 10
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match           34.1%;  Score 2227.5;  DB 2;  Length 1291;
  Best Local Similarity  37.8%;  Pred. No. 1.6e-78;
  Matches  495;  Conservative  273;  Mismatches  456;  Indels   87;  Gaps   33;

Qy          4 INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 60
              ||:|||||:::  |: ::|    |   :||:| |   |||||||   |  |:||:  :
Db          5 INNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNKSSG 64

Qy         61 LKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNT 119
             : | |   ||||:|| ::::|: ||: : |:|||| :   |  ||| :    ||||:
Db         65 IFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLGDRRV 124

Qy        120 PDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNNYMPS 174
              |  :|:   ||    |  :   ::     :  |:|| |  | | |  : :| ::|::  |
Db        125 PLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHF-AS 183

Qy        175 NHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGLYGAK 225
              |||  |  : | |||   ||:   |:          :  |||| ||||||| ||||||  |
Db        184 REGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLYGIK 243

Qy        226 GITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASK 285
              :     :  ::   : :      ||  |||| | :||: :    ||  :| ::: |  :
Db        244 -VNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIVDR 302

Qy        286 LSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFTEFDLA 340
              |:|| |      |  :| ||: |: ||   :|: | ||:::   |: ::| | : ||| ::|
Db        303 LNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETNIA 362

Qy        341 TKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNPRIIT 397
              :::| | :|        |: |||:: || | ||:||:  |::  :|||| :| :
Db        363 ENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQAYE 422

Qy        398 PITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT 457
              |:   |   |: ||   |||    |||:::| :|||:| :||::|| ::   : |
Db        423 EISKEHLAVYKIQMCK---SVKA--PGICIDVDNEDLFFIADKNSFSDD-LSKNERIAYN 476
```

```
Qy        458 VTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVN 515
Db        477 -TQNNYIDNDFSINELILDTDLISKIELPSENTESLTDFNVYVPEYKKQPA--IKKIFTD 533

Qy        516 ELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSW 575
Db        534 ENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGW 593

Qy        576 IQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLE 635
Db        594 VKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLE 653

Qy        636 FEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINT 695
Db        654 FIPELLIPVVGAFLLESYI---DNKNKIIETINSALTKRDEKWIDMYGLIVAQWLSTVNT 710

Qy        696 QFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAM 755
Db        711 QFYTIKEGMYKALNYQAQALEEIIKYKYNIYSEKERSNI--NIDFNDVNSKLNEGINQAI 768

Qy        756 NNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMV 815
Db        769 DNINNFINECSVSYLMKKMIPLAVEKLLDFDNTLRKNLLNYIDENKLYLIGSAEYEKSKV 828

Qy        816 TDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININ 875
Db        829 DKYLKTSIPFDLSTYTNNTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVY 888

Qy        876 GDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--V 933
Db        889 DGVKL--NDKNQFKLTSSANSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYI 946

Qy        934 NNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFV 993
Db        947 HNEYTIINCMK-NNSGWKISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFV 1005

Qy        994 TITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDK 1052
Db       1006 TITNNS-DNAKIYINGKLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNT 1064

Qy       1053 ELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDS------TL 1106
Db       1065 ELSQSNIEEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTR 1124

Qy       1107 SINNIRSTILLANRLYSGIKVKIQRVNNS-STNDNLVRKNDQVYINFV----ASKTHLFP 1161
Db       1125 SKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYIYK 1184

Qy       1162 LYADTATTNKEKTI--KISSSGNRFNQVVVM---NSVGNNCTMNFKNN--NGNNIGLLG- 1213
Db       1185 YF----KKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGI 1240

Qy       1214 ---------FK--ADTVVASTWYYTHM--RDHTNSNGCFWNFISEEHGWQE 1251
Db       1241 HRFYESGIVFKEYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291
```

RESULT 11
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR PIR; I40631; I40631.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 33.8%; Score 2208; DB 2; Length 1291;
Best Local Similarity 37.6%; Pred. No. 9.3e-78;
Matches 496; Conservative 272; Mismatches 454; Indels 98; Gaps 33;

```
Qy          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56
Db          1 MPVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFFQTLIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        116 NDNTPDNQFHIGDASAVEIKFSNGSQDI-----LLPNVIIMGAEPDLFETNSSNISLRNN 170
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221
Db        181 F-ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGL 239

Qy        222 YGAKGITTKYTITQKQNPLITN-----IRGTN---IEEFLTFGGTDLNIITSAQSNDIYT 273
Db        240 YGIK---------VDDLPIVPNEKKFFMQSTDTIQAEELYTFGGQDPSIISPSTDKSIYD 290

Qy        274 NLLADYKKIASKLSKVQV----SNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFK 329
Db        291 KVLQNFRGIVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFNKLYK 350

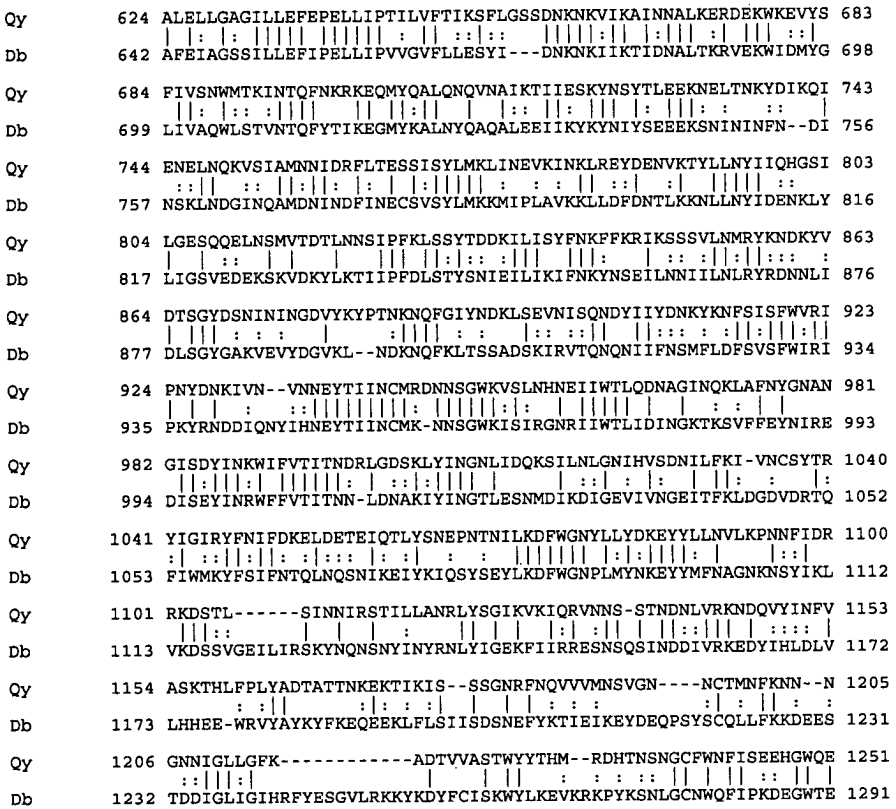
Qy        330 KL-YSFTEFDLATKFQVKCRQTYIG-QYKYFKLSNLLNDSIYNISEGYNIN--NLKVNFR 385
Db        351 SLMLGFTEINIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMGKEYR 410

Qy        386 GQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYND 445
Db        411 GQNKAINKQAYEEISKEHLAVYKIQMCKS-VKVPG----ICIDVDNENLFFIADKNSFSD 465

Qy        446 DNINTPKEIDDTVTSNNNYEND--LDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDS 503
Db        466 D--LSKNERVEYNTQNNYIGNDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEK 523

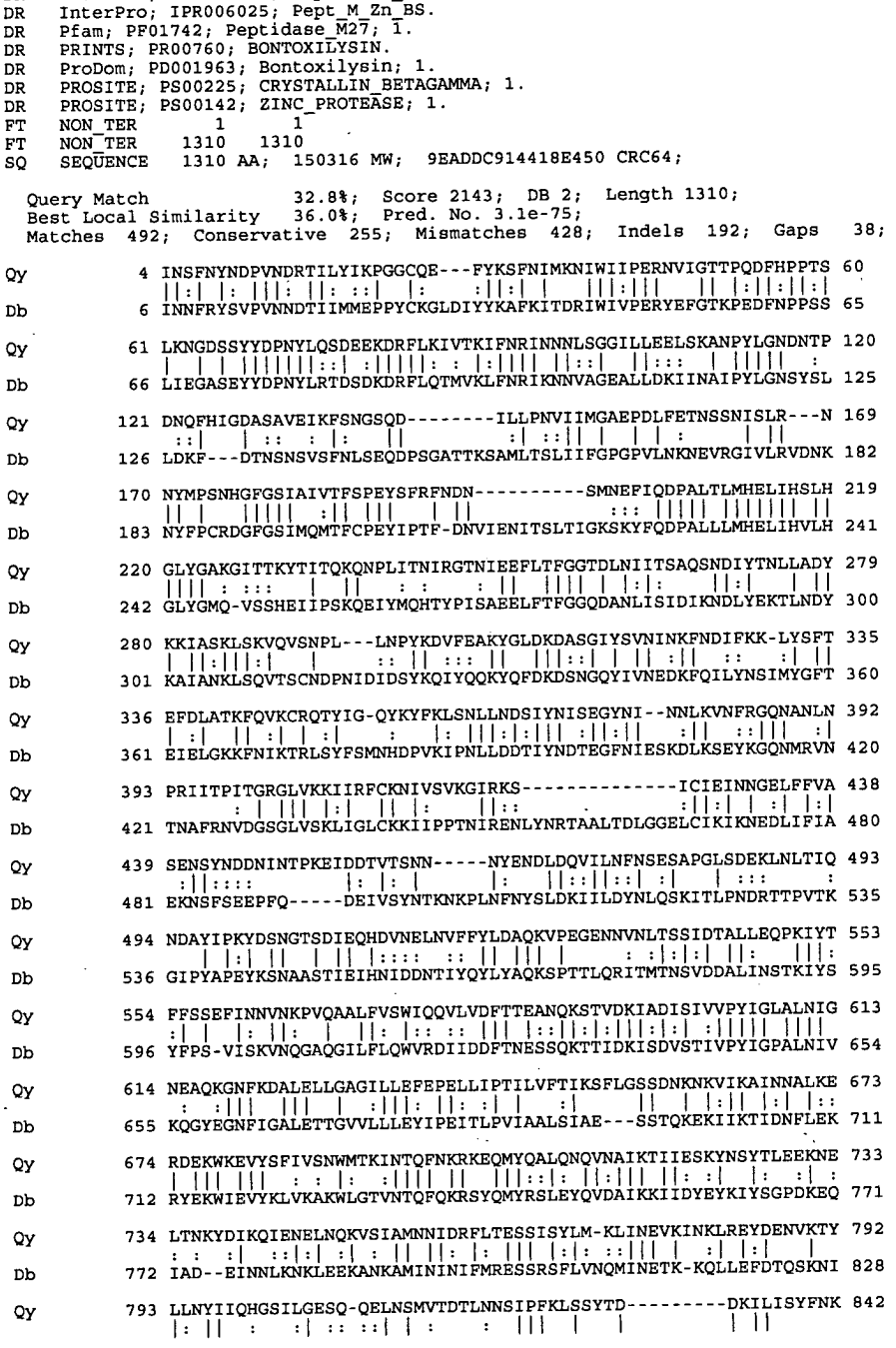
Qy        504 NGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNV 563
Db        524 QPA--IKKVFTDENTIFQYLYSQTFPLNIRDISLTSSFDDALLVSSKVYSFFSMDYIKTA 581

Qy        564 NKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKD 623
Db        582 NKVVEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGDETAKGNFES 641
```



```
RESULT 12
Q93N27
ID   Q93N27      PRELIMINARY;      PRT;  1310 AA.
AC   Q93N27;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tetanus toxin (Fragment).
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shumin Z., Dianliang L.;
RT   "Cloning and sequence analysis of tetanus toxin gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF389424; AAK72964.2; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER    1310   1310
SQ   SEQUENCE   1310 AA;  150316 MW;  9EADDC914418E450 CRC64;

  Query Match             32.8%;  Score 2143;  DB 2;  Length 1310;
  Best Local Similarity   36.0%;  Pred. No. 3.1e-75;
  Matches  492;  Conservative  255;  Mismatches  428;  Indels  192;  Gaps   38;
```



```
Db        829 LMQYIKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVIL------ 882

Qy        843 FFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP-TNKNQFGIYNDKLSEVNIS 901
              | |::||:   ||   | ||::|::    |    |     : |:: ||| :
Db        883 -----KKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVH 937

Qy        902 QNDYIIYDNKYKNFSISFWVRIP--------NYDNKIVNVNNEYTIINCMRDNN----SG 949
              :   | |:: : ||::|||:|:|        ||       |||:||: |:  :    ||
Db        938 KAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYD------TNEYSIISSMKKYSLSIGSG 991

Qy        950 WKVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISD----YI-NKWIFVTITNDRLGDSK 1004
              | |||  | :||||:|:||  ::: |       :||    |: |||:|:|||||||  :
Db        992 WSVSLKGNNLIWTLKDSAGEVRQITF-----RDLSDKFNAYLANKWVFITITNDRLSSAN 1046

Qy       1005 LYINGNLIDQKSILNLGNIHVSDNILFKIVNC-SYTRYIGIRYFNIFDKELDETEIQTLY 1063
              ||||| |:    |  || |   :||  |:  | :  :|: |  | || | |:  ||: ||
Db       1047 LYINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLY 1106

Qy       1064 SNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLAN---- 1119
              ::  :   |:||||| | || ||||: |  :      ||  : : ||   : | |
Db       1107 TSYLSITFLRDFWGNPLRYDTEYYLIPVAYSS------KD--VQLKNITDYMYLTNAPSY 1158

Qy       1120 ----------RLYSGIKVKIQRVNNSSTNDNLVRKND--QVYINFVASKTHLFPLYADTA 1167
                      |||||:|  |:|   ::  |: ||  |  ::|:::
Db       1159 TNGKLNIYYRRLYSGLKFIIKRYTPNNEIDSFVRSGDFIKLYVSY-------------- 1203

Qy       1168 TTNKEKTIKISSSGNRFNQVVVMNSVGNNC---------------------TMNFKNNN 1205
              | |  :     || || :  :  || |                        :   ::
Db       1204 -NNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDK 1262

Qy       1206 GNNIGLLGF---------KADTVVASTWYYTHMRDHTNSNGCFWNFI 1243
                ::||:|           | ::|| ||: |::| | :  | | |:
Db       1263 DASLGLVGTHNGQIGNDPNRDILIASNWYFNHLKDKTLT--CDWYFV 1307
```

RESULT 13
Q45862
ID Q45862 PRELIMINARY; PRT; 367 AA.
AC Q45862;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Botulinum neurotoxin type E (Fragment).
GN BONT/E.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=type E;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70815; CAA50146.1; -.
DR PIR; S21178; S21178.
DR HSSP; P10845; 3BTA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 42854 MW; 0810595B3A865570 CRC64;

Query Match 29.1%; Score 1904; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.3e-66;
Matches 367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180

Qy        796 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
              |||||||
Db        361 NYGNANG 367
```

RESULT 14
Q45861
ID Q45861 PRELIMINARY; PRT; 367 AA.
AC Q45861;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Botulinum neurotoxin type E (Fragment).
GN BONT/E.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=type E;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70818; CAA50149.1; -.
DR PIR; S48106; S48106.
DR HSSP; P10845; 3BTA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 42902 MW; 346A610C2FF70262 CRC64;

Query Match 29.0%; Score 1895; DB 2; Length 367;
Best Local Similarity 99.7%; Pred. No. 3e-66;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        616 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERD 60

Qy        676 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELT 120

Qy        736 NKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NKYDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLN 180
```

```
Qy        796 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNM 240

Qy        856 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNF 300

Qy        916 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        976 NYGNANG 982
              |||||||
Db        361 NYGNANG 367
```

RESULT 15
Q9QTG7
ID Q9QTG7 PRELIMINARY; PRT; 1275 AA.
AC Q9QTG7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NTX (Fragment).
GN NTX.
OS Clostridium botulinum D bacteriophage.
OC Viruses.
OX NCBI_TaxID=29342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1873;
RX MEDLINE=99017546; PubMed=9802560;
RA Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT "Molecular composition of the 16S toxin produced by a Clostridium
RT botulinum type D strain, 1873.";
RL Microbiol. Immunol. 42:599-605(1998).
DR EMBL; AB012112; BAA75084.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1275 1275
SQ SEQUENCE 1275 AA; 146742 MW; 3C50F46C8233E2D6 CRC64;

  Query Match 28.7%; Score 1880; DB 12; Length 1275;
  Best Local Similarity 34.3%; Pred. No. 4.6e-65;
  Matches 455; Conservative 250; Mismatches 486; Indels 134; Gaps 41;

```
Qy          4 INSFNYNDPVNDRTILYIK---------PGGCQEFYKSFNIMKNIWIIPERNVIGTTPQD 54
              :  |||:|||||  |||::         |       |:| | :|||:||||    | |
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTP------VKAFMITQNIWVIPERFSSDTNPSL 58

Qy         55 FHPPTSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114
                ||       |||||:|| :||:|| ||| : |:| |||    |  |:  |   :|::
Db         59 SKPPRPTSK-YQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFM 117

Qy        115 GNDNTPDNQFHIGDAS---AVEIKFSNGS---QDILLPNVIIMGAEPDLFETNSSNISLR 168
              |: :||:: |     :  ||| || |||    :|: |:|:| |  |:: :  :|
Db        118 GDSSTPEDTFDFTRHTTNIAVE-KFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQG 176

Qy        169 NNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLH 219
                 ||  |||:::|:  :||:   |:| : |:          |  || : ||||| ||||
Db        177 QQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLH 236

Qy        220 GLYGAKGITTKYTITQKQNPLITNIRGTNI--EEFLTFGGTDLNIITSAQSNDIYTNLLA 277
              |||  | :  | :  :   :  | |:  ||  |||| |: ||   : : :    |
Db        237 QLYGI-NIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALG 295

Qy        278 DYKKIASKLSKVQVSNPL-----LNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLY 332
              || || :|: :  : |       :: || :|  ||  ||| :| : |||:||| ::  |
Db        296 HYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLT 355

Qy        333 S-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFR--GQNA 389
              :  :|   :::: || |  |  ::     :|:|:|:|| | :|:|: |    |   |||
Db        356 NVMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNI 415

Qy        390 NLNPRIITPITGRGLVKKIIRFCKNIVSVKGIR-KSICIEINNGELFFVASENSYNDDNI 448
              ||  :  ::   :|    : |  :   |  |  | ||:: |  | :|| ::| : :
Db        416 ERNP-ALQKLSSESVVDLFTKVCLRL--TKNSRDDSTCIKVKNNRLPYVADKDSISQEIF 472

Qy        449 NTPKEIDDTVTSNNNYENDLDQVILN----FNSESA----PGLSDEKLNLTIQNDAYIPK 500
                  |:|   | : :  ||: ||:     | |        | :: | |||  :   :
Db        473 ENKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVF--- 529

Qy        501 YDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFI 560
              ||      || :: |: || ::||::||:      |: ||:|:: ||     |||||  |
Db        530 YD-----DITKY-VDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPS-LA 582

Qy        561 NNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGN 620
              ||| ||| ||::|  :|: ||||   :| |:|||:|:|:::|||| |||||| | :||
Db        583 EKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGN 642

Qy        621 FKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKE 680
              |  |   |  ||| ||  || : |||   |  |   : |:|| | | |::| ::||:
Db        643 FNQAFATAGVAFLLEGFPEFTIPALGVFT---FYSSIQEREKIIKTIENCLEQRVKRWKD 699

Qy        681 VYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDI 740
              | ::|||:::| ||||.  ||| :|  | :|||  |: :|  |:  :|  :  |  :
Db        700 SYQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENI--KSQV 757

Qy        741 KQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQH 800
              : ::| |: |:| |||||::|: | |::|| | :    |::| ::|   || |:| | |
Db        758 ENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSH 817

Qy        801 GSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 860
              ||     | : | ::  |::|| : |||:: :|    |::|  |  | :|::: | :
Db        818 NIILVGEVDRLKAKVNESFENTMPFNIPSYTNNSLLKDIINEYFNSINDSKILSLQNKKN 877

Qy        861 KYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS----EVNISQNDYIIYDNKYKNFS 916
              ||||||:: : :  :|      | |     |||    :: :: |: |:|    |:| |
Db        878 ALVDTSGYNAEVRVGDNVQLNTIYTNDF-----KLSSSGDKIIVNLNNNILYSAIYENSS 932

Qy        917 ISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQK---L 973
              :|||::|      : | :|||||| : : |||||: : :  | | |||   :|:|   |
Db        933 VSFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQD---VNRKYKSL 985

Qy        974 AFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI 1033
              |:|  :   : | ||| ||||||: :|  |||||  |   : | :|  : :    |:| |
Db        986 IFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGI 1045

Qy       1034 -VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVL 1092
               |    : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|
Db       1046 DENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN-- 1103

Qy       1093 KPNNFIDRRKDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSS------TNDN 1140
                :|:|||             | :|:       ::||:|  : |: |:: :        ||
Db       1104 --DNYIDRYIAP-------ESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDN 1154

Qy       1141 LVRK---NDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVV--VMNSVGN 1195
              ::     | : |:  :     ::       :  |    :|: |:   :   : | |
Db       1155 IILHMLYNSRKYM-IIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSK 1213

Qy       1196 N--CTMNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDHTNSNGCFWNFIS 1244
```

```
Db       1214 NKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFIS 1269

Qy       1245 EEHGW 1249
                : ||
Db       1270 RDPGW 1274
```

Search completed: September 10, 2004, 03:41:37
Job time : 61.2852 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 75.642 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 1291
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1058	82.0	1291	2	AAY05814	Aay05814 Non-toxic
2	449	34.8	450	3	AAY77136	Aay77136 Synthetic
3	449	34.8	450	4	AAB04092	Aab04092 Botulism
4	444	34.4	444	5	ABG69069	Abg69069 Botulinum
5	436	33.8	436	5	ABG69078	Abg69078 Botulinum
6	436	33.8	462	2	AAW68397	Aaw68397 Clostridi
7	436	33.8	811	5	ABG69079	Abg69079 Botulinum
8	398	30.8	399	4	AAB04098	Aab04098 Botulism
9	68	5.3	386	4	AAB04099	Aab04099 Botulism
10	68	5.3	824	5	ABG69081	Abg69081 Botulinum
11	52	4.0	399	3	AAY78982	Aay78982 C. botuli
12	22	1.7	451	4	AAB04093	Aab04093 Botulism
13	11	0.9	207	4	AAG79295	Aag79295 Amino aci
14	11	0.9	407	3	AAY77141	Aay77141 Native bo
15	11	0.9	413	4	AAB04168	Aab04168 Botulism
16	11	0.9	413	5	ABG69074	Abg69074 Botulinum
17	11	0.9	423	3	AAB36302	Aab36302 C. botuli
18	11	0.9	451	2	AAW68398	Aaw68398 Clostridi
19	11	0.9	837	3	AAY77140	Aay77140 Native bo
20	11	0.9	847	4	AAB04081	Aab04081 Botulism
21	11	0.9	861	5	ABG69075	Abg69075 Botulinum
22	11	0.9	871	2	AAW56008	Aaw56008 Botulinum
23	11	0.9	871	2	AAW56019	Aaw56019 Recombina
24	11	0.9	871	2	AAW56007	Aaw56007 Recombina
25	11	0.9	873	2	AAW56016	Aaw56016 Recombina
26	11	0.9	875	2	AAW56009	Aaw56009 Recombina
27	11	0.9	878	2	AAW56010	Aaw56010 Recombina
28	11	0.9	894	2	AAW56015	Aaw56015 Recombina
29	11	0.9	907	2	AAW56012	Aaw56012 Recombina
30	11	0.9	953	2	AAW56011	Aaw56011 Recombina
31	11	0.9	1013	2	AAW56013	Aaw56013 Recombina
32	11	0.9	1067	3	AAY93307	Aay93307 A mangane
33	11	0.9	1092	3	AAY93310	Aay93310 A mangane
34	11	0.9	1295	5	AAU99339	Aau99339 Clostridi
35	11	0.9	1296	2	AAR95010	Aar95010 C. botuli
36	10	0.8	410	4	AAB04102	Aab04102 Botulism
37	10	0.8	848	5	ABG69087	Abg69087 Botulinum
38	9	0.7	16	2	AAW05608	Aaw05608 Tetanus t
39	9	0.7	20	3	AAY96457	Aay96457 Tetanus t
40	9	0.7	26	5	AAU10838	Aau10838 Human cyt
41	9	0.7	27	5	ABB79188	Abb79188 Human cyt
42	9	0.7	28	5	ABB79185	Abb79185 Human cyt
43	9	0.7	28	5	ABB79186	Abb79186 Human cyt
44	9	0.7	28	5	ABB79187	Abb79187 Human cyt
45	9	0.7	30	2	AAR62711	Aar62711 LHRH-cont

ALIGNMENTS

RESULT 1
AAY05814
ID AAY05814 standard; protein; 1291 AA.
XX
AC AAY05814;
XX
DT 02-AUG-1999 (first entry)
XX
DE Non-toxic modified botulinum toxin rBoNT/C.
XX
KW Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery; mutant.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 229
FT /note= "His in native toxin"
FT Misc-difference 230
FT /note= "Glu in native toxin"
FT Misc-difference 233
FT /note= "His in native toxin"
XX
PN WO9920306-A1.
XX
PD 29-APR-1999.
XX
PF 16-OCT-1998; 98WO-US021897.
XX
PR 20-OCT-1997; 97US-00954302.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Simpson L, Kiyatkin N, Maksymowych A;
XX
DR WPI; 1999-302646/25.
DR N-PSDB; AAX25521.
XX
PT Modified toxin useful for systemic delivery of oral vaccines and
PT therapeutic agents.
XX
PS Example 1; Page 31-33; 37pp; English.
XX
CC The present sequence represents a modified serotype C botulinum toxin,
CC termed rBoNT/C, in which amino acids His-229, Glu-230 and His-233 of the
CC native sequence are substituted by Gly, Thr and Asn, respectively, i.e.
CC the zinc binding motif (see AAY05817) of the light chain holotoxin is
CC modified, resulting in loss of endoprotease activity. DNA coding for the

CC modified botulinum toxin (see AAX25521) was assembled from 3 separate
CC toxin fragments using PCR and site-directed mutagenesis. The modified
CC recombinant botulinum toxin maintains its ability to translocate from the
CC gut into the general circulation but is non-toxic. It can be used as an
CC oral vaccine for antigenic peptides including botulinum toxin (i.e. an
CC oral vaccine for botulism) or for the oral delivery of other therapeutic
CC agents to the general circulation
XX
SQ Sequence 1291 AA;

Query Match 82.0%; Score 1058; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1058; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       234 AMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKA 293
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       234 AMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKA 293

Qy       294 LDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNE 353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       294 LDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNE 353

Qy       354 LTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQ 413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       354 LTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQ 413

Qy       414 NLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVK 473
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       414 NLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVK 473

Qy       474 TDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVF 533
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       474 TDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVF 533

Qy       534 YDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGV 593
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       534 YDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGV 593

Qy       594 QGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAV 653
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       594 QGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAV 653

Qy       654 TGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWL 713
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       654 TGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWL 713

Qy       714 SRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKIS 773
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       714 SRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKIS 773

Qy       774 EAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLK 833
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       774 EAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLK 833

Qy       834 AKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEV 893
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       834 AKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEV 893

Qy       894 SEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNL 953
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       894 SEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNL 953

Qy       954 PGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVT 1013
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       954 PGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVT 1013

Qy      1014 NNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDF 1073
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1014 NNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDF 1073

Qy      1074 YIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVF 1133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1074 YIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVF 1133

Qy      1134 NTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 1193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1134 NTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 1193

Qy      1194 STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR 1253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1194 STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR 1253

Qy      1254 LGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
            ||||||||||||||||||||||||||||||||||||||
Db      1254 LGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 2
AAY77136
ID AAY77136 standard; protein; 450 AA.
XX
AC AAY77136;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype C; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87214.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 24; Page 41-42; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences

```
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 450 AA;

  Query Match             34.8%;  Score 449;  DB 3;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  449;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        843 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 61

Qy        903 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

Qy        963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 181

Qy       1023 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241

Qy       1083 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 301

Qy       1143 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361

Qy       1203 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 421

Qy       1263 YLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db        422 YLVPTVKQGNYASLLESTSTHWGFVPVSE 450


RESULT 3
AAB04092
ID   AAB04092 standard; protein; 450 AA.
XX
AC   AAB04092;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype C).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54486.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 5b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 450 AA;

  Query Match             34.8%;  Score 449;  DB 4;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  449;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        843 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 61

Qy        903 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

Qy        963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 181

Qy       1023 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241

Qy       1083 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 301

Qy       1143 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361

Qy       1203 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 421

Qy       1263 YLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db        422 YLVPTVKQGNYASLLESTSTHWGFVPVSE 450


RESULT 4
ABG69069
ID   ABG69069 standard; protein; 444 AA.
XX
AC   ABG69069;
XX
```

```
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #3.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98539.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 121-122; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 444 AA;

  Query Match             34.4%;  Score 444;  DB 5;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  444;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGR 444
              ||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGR 444


RESULT 5
ABG69078
ID   ABG69078 standard; protein; 436 AA.
XX
AC   ABG69078;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #12.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98548.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 140-141; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 436 AA;
```

```
  Query Match             33.8%;  Score 436;  DB 5;  Length 436;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 60

Qy         62 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 120

Qy        122 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 180

Qy        182 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 240

Qy        242 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 300

Qy        302 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 360

Qy        362 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 420

Qy        422 RKVNPENMLYLFTKFC 437
              ||||||||||||||||
Db        421 RKVNPENMLYLFTKFC 436
```

```
RESULT 6
AAW68397
ID   AAW68397 standard; protein; 462 AA.
XX
AC   AAW68397;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type C1 toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW   botulism; BotC.
XX
OS   Clostridium botulinum; serotype C1 strain Stockholm.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30588.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 45; Page 339-341; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC   DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC   express BotC soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 462 AA;
```

```
  Query Match             33.8%;  Score 436;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86

Qy        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146

Qy        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206

Qy       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266

Qy       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326

Qy       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386

Qy       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446

Qy       1276 LLESTSTHWGFVPVSE 1291
              ||||||||||||||||
Db        447 LLESTSTHWGFVPVSE 462
```

```
RESULT 7
ABG69079
ID   ABG69079 standard; protein; 811 AA.
XX
AC   ABG69079;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #13.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
```

KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98549.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 142-144; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 811 AA;

Query Match 33.8%; Score 436; DB 5; Length 811;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 60

Qy         62 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 120

Qy        122 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 180

Qy        182 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 240

Qy        242 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 300

Qy        302 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 360

Qy        362 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 420

Qy        422 RKVNPENMLYLFTKFC 437
              ||||||||||||||||
Db        421 RKVNPENMLYLFTKFC 436
```

RESULT 8
AAB04098
ID AAB04098 standard; protein; 399 AA.
XX
AC AAB04098;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain N-terminal sequence (serotype C).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54494.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 6; Fig 13b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 399 AA;

Query Match 30.8%; Score 398; DB 4; Length 399;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        445 SLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSK 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSK 61

Qy        505 NTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTF 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 NTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTF 121

Qy        565 TRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 181

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 241

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 301

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 361

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQN 842
              ||||||||||||||||||||||||||||||||||||||
Db        362 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQN 399
```

RESULT 9
AAB04099
ID AAB04099 standard; protein; 386 AA.
XX
AC AAB04099;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain N-terminal sequence (serotype D).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54495.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 6; Fig 14b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 386 AA;

```
  Query Match             5.3%;  Score 68;  DB 4;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-62;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 352

Qy        797 PKVIDELN 804
              ||||||||
Db        353 PKVIDELN 360
```

RESULT 10
ABG69081
ID ABG69081 standard; protein; 824 AA.
XX
AC ABG69081;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #15.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98551.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 147-149; 166pp; English.

```
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 824 AA;

  Query Match             5.3%;  Score 68;  DB 5;  Length 824;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-61;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        731 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 790

Qy        797 PKVIDELN 804
              ||||||||
Db        791 PKVIDELN 798


RESULT 11
AAY78982
ID   AAY78982 standard; protein; 399 AA.
XX
AC   AAY78982;
XX
DT   20-JUN-2000  (first entry)
XX
DE   C. botulinum type D toxin amino acid sequence.
XX
KW   Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
KW   protect.
XX
OS   Clostridium botulinum.
XX
PN   WO200005252-A1..
XX
PD   03-FEB-2000.
XX
PF   20-JUL-1999;   99WO-IB001301.
XX
PR   22-JUL-1998;   98ZA-00006538.
XX
PA   (AGRI-) AGRIC RES COUNCIL.
XX
PI   De Bruyn EE,  Botha AD;
XX
DR   WPI; 2000-205375/18.
DR   N-PSDB; AAZ98630, AAZ98631.
XX
PT   Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin type
PT   D, useful in vaccines for protection against botulism, comprises at least
PT   one amino acid mutation not present in the wild type D neurotoxins.
XX
PS   Claim 3, 4; Page 54-57; 66pp; English.
XX
CC   This sequence represents the amino acid sequence of a synthetic non-toxic
CC   immunogenic derivative of Clostridium botulinum type D toxin (BoNT).
CC   Botulinum neurotoxin causes botulism poisoning in cattle and sheep, and
CC   usually results in the death of the affected or poisoned animal. The non-
CC   toxic immunogenic fragments of the C. botulinum neurotoxin are useful in
CC   vaccines to protect animals (e.g. humans, cattle, sheep, pigs) against
CC   BoNT type D poisoning. The non-toxic fragments can be produced relatively
CC   simply and inexpensively (specifically by fermentation techniques). As
CC   the fragments are not toxic the risk to production staff is reduced
XX
SQ   Sequence 399 AA;

  Query Match             4.0%;  Score 52;  DB 3;  Length 399;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-45;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         44 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 95


RESULT 12
AAB04093
ID   AAB04093 standard; protein; 451 AA.
XX
AC   AAB04093;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype D).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54487.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 6b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
```

```
XX
SQ   Sequence 451 AA;

  Query Match             1.7%;  Score 22;  DB 4;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        845 PFNIFSYTNNSLLKDIINEYFN 866
              ||||||||||||||||||||||
Db         16 PFNIFSYTNNSLLKDIINEYFN 37


RESULT 13
AAG79295
ID   AAG79295 standard; protein; 207 AA.
XX
AC   AAG79295;
XX
DT   03-JAN-2002  (first entry)
XX
DE   Amino acid sequence of botulinum neurotoxin (BoNT) fragment.
XX
KW   Neurotoxin; BoNT; vaccine; botulism.
XX
OS   Clostridium botulinum.
XX
PN   US6287566-B1.
XX
PD   11-SEP-2001.
XX
PF   19-MAY-1995;   95US-00446114.
XX
PR   19-MAY-1995;   95US-00446114.
XX
PA   (USSA ) US SEC OF ARMY.
XX
PI   Dertzbaugh MT;
XX
DR   WPI; 2001-615462/71.
XX
PT   New protective epitopes of neurotoxin of Clostridium botulinum, useful
PT   for developing vaccines against neurotoxins of Clostridium botulinum and
PT   for identifying protective antibodies.
XX
PS   Claim 1; Col 25; 14pp; English.
XX
CC   The present sequence represents a fragment of the Clostridium botulinum
CC   neurotoxin (BoNT). It was produced by amplifying overlapping fragments of
CC   the BoNT gene. The amplified fragments were cloned expressed to identify
CC   immunogenic polypeptides which are capable of giving rise to protective
CC   antibodies. The BoNT polypeptide fragment are useful as vaccines, for
CC   immunizing against botulism, and as diagnostic agents to identify
CC   protective antibodies
XX
SQ   Sequence 207 AA;

  Query Match             0.9%;  Score 11;  DB 4;  Length 207;
  Best Local Similarity   100.0%;  Pred. No. 0.029;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||||
Db        179 IIPYIGPALNI 189


RESULT 14
AAY77141
ID   AAY77141 standard; protein; 407 AA.
XX
AC   AAY77141;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Native botulinum neurotoxin serotype A (BoNTA) N-terminal fragment (Hn).
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype A; N-terminal fragment;
KW   Hn; Venezuelan equine encephalitis virus replicon; VEE; botulism;
KW   vaccine; diagnosis; drug screening.
XX
OS   Clostridium botulinum.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87219.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Example 3; Page 51; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. The present sequence represents the native BoNTA heavy
CC   chain N-terminal fragment (Hn) used in an exemplification of the present
CC   invention
XX
SQ   Sequence 407 AA;

  Query Match             0.9%;  Score 11;  DB 3;  Length 407;
  Best Local Similarity   100.0%;  Pred. No. 0.053;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||||
Db        180 IIPYIGPALNI 190


RESULT 15
AAB04168
ID   AAB04168 standard; protein; 413 AA.
XX
AC   AAB04168;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain N-terminal sequence (serotype A).
XX
```

KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54590.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Fig 11b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 413 AA;

Query Match 0.9%; Score 11; DB 4; Length 413;
Best Local Similarity 100.0%; Pred. No. 0.054;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALNI 638
||||||||||||
Db 186 IIPYIGPALNI 196

Search completed: September 10, 2004, 05:05:59
Job time : 77.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 21.1666 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 1291
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	0.9	207	3	US-08-446-114A-21	Sequence 21, Appl
2	11	0.9	871	4	US-09-255-829-2	Sequence 2, Appli
3	11	0.9	871	4	US-09-255-829-8	Sequence 8, Appli
4	11	0.9	871	4	US-09-255-829-26	Sequence 26, Appl
5	11	0.9	873	4	US-09-255-829-6	Sequence 6, Appli
6	11	0.9	875	4	US-09-255-829-10	Sequence 10, Appl
7	11	0.9	878	4	US-09-255-829-12	Sequence 12, Appl
8	11	0.9	894	4	US-09-255-829-4	Sequence 4, Appli
9	11	0.9	907	4	US-09-255-829-16	Sequence 16, Appl
10	11	0.9	953	4	US-09-255-829-14	Sequence 14, Appl
11	11	0.9	1013	4	US-09-255-829-18	Sequence 18, Appl
12	11	0.9	1296	1	US-08-480-604A-28	Sequence 28, Appl
13	11	0.9	1296	2	US-08-405-496A-28	Sequence 28, Appl
14	11	0.9	1296	3	US-08-915-136-28	Sequence 28, Appl
15	11	0.9	1296	4	US-09-084-517-28	Sequence 28, Appl
16	9	0.7	16	1	US-08-446-692-44	Sequence 44, Appl
17	9	0.7	16	2	US-08-488-351A-44	Sequence 44, Appl
18	9	0.7	16	5	PCT-US95-13841-16	Sequence 16, Appl
19	9	0.7	26	4	US-09-692-170C-19	Sequence 19, Appl
20	9	0.7	27	4	US-09-983-019-14	Sequence 14, Appl
21	9	0.7	28	4	US-09-983-019-11	Sequence 11, Appl
22	9	0.7	28	4	US-09-983-019-12	Sequence 12, Appl
23	9	0.7	28	4	US-09-983-019-13	Sequence 13, Appl
24	9	0.7	30	1	US-08-446-692-23	Sequence 23, Appl
25	9	0.7	30	2	US-08-488-351A-23	Sequence 23, Appl
26	9	0.7	38	1	US-08-446-692-59	Sequence 59, Appl
27	9	0.7	38	1	US-08-446-692-65	Sequence 65, Appl
28	9	0.7	38	2	US-08-488-351A-59	Sequence 59, Appl
29	9	0.7	38	2	US-08-488-351A-65	Sequence 65, Appl
30	9	0.7	48	1	US-08-446-692-37	Sequence 37, Appl
31	9	0.7	48	2	US-08-488-351A-37	Sequence 37, Appl
32	9	0.7	853	4	US-08-913-880C-17	Sequence 17, Appl
33	9	0.7	858	4	US-08-913-880C-16	Sequence 16, Appl
34	9	0.7	860	4	US-08-913-880C-15	Sequence 15, Appl
35	9	0.7	862	4	US-08-913-880C-14	Sequence 14, Appl
36	9	0.7	865	4	US-08-913-880C-13	Sequence 13, Appl
37	9	0.7	866	4	US-08-913-880C-12	Sequence 12, Appl
38	9	0.7	874	4	US-08-913-880C-11	Sequence 11, Appl
39	9	0.7	875	4	US-08-913-880C-10	Sequence 10, Appl
40	9	0.7	1315	4	US-08-913-880C-1	Sequence 1, Appli
41	8	0.6	412	2	US-08-978-458-2	Sequence 2, Appli
42	8	0.6	438	1	US-08-480-604A-23	Sequence 23, Appl
43	8	0.6	438	2	US-08-405-496A-23	Sequence 23, Appl
44	8	0.6	438	3	US-08-915-136-23	Sequence 23, Appl
45	8	0.6	438	4	US-09-084-517-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-446-114A-21
; Sequence 21, Application US/08446114A
; Patent No. 6287566
; GENERAL INFORMATION:
; APPLICANT: Dertzbaugh, Mark T.
; TITLE OF INVENTION: PROTECTIVE PEPTIDES OF NEUROTOXIN OF C.
; TITLE OF INVENTION: BOTULINUM
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Associates
; STREET: 9669 A Main Street
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,114A
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: der08/446,114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-425-4250
; TELEFAX: 703-425-2767
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: C. botulinum
US-08-446-114A-21

Query Match 0.9%; Score 11; DB 3; Length 207;
Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        628 IIPYIGPALNI 638
              ||||||||||
Db        179 IIPYIGPALNI 189


RESULT 2
US-09-255-829-2
; Sequence 2, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-2

  Query Match             0.9%;  Score 11;  DB 4;  Length 871;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||
Db        633 IIPYIGPALNI 643


RESULT 3
US-09-255-829-8
; Sequence 8, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-8

  Query Match             0.9%;  Score 11;  DB 4;  Length 871;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||
Db        633 IIPYIGPALNI 643


RESULT 4
US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-26

  Query Match             0.9%;  Score 11;  DB 4;  Length 871;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||| 
Db        633 IIPYIGPALNI 643


RESULT 5
US-09-255-829-6
; Sequence 6, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  873 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-6

  Query Match             0.9%;  Score 11;  DB 4;  Length 873;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||| 
Db        635 IIPYIGPALNI 645


RESULT 6
US-09-255-829-10
; Sequence 10, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  875 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-10

  Query Match             0.9%;  Score 11;  DB 4;  Length 875;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||| 
Db        637 IIPYIGPALNI 647


RESULT 7
US-09-255-829-12
; Sequence 12, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
```

; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-12

Query Match 0.9%; Score 11; DB 4; Length 878;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALNI 638
||||||||||||
Db 640 IIPYIGPALNI 650

RESULT 8
US-09-255-829-4
; Sequence 4, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-4

Query Match 0.9%; Score 11; DB 4; Length 894;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALNI 638
||||||||||||
Db 656 IIPYIGPALNI 666

RESULT 9
US-09-255-829-16
; Sequence 16, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-16

Query Match 0.9%; Score 11; DB 4; Length 907;
Best Local Similarity 100.0%; Pred. No. 0.022;

```
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||||
Db        640 IIPYIGPALNI 650


RESULT 10
US-09-255-829-14
; Sequence 14, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  953 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-14

  Query Match             0.9%;  Score 11;  DB 4;  Length 953;
  Best Local Similarity   100.0%;  Pred. No. 0.023;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||||
Db        640 IIPYIGPALNI 650


RESULT 11
US-09-255-829-18
; Sequence 18, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1013 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-18

  Query Match             0.9%;  Score 11;  DB 4;  Length 1013;
  Best Local Similarity   100.0%;  Pred. No. 0.024;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||||
Db        640 IIPYIGPALNI 650


RESULT 12
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-28

Query Match 0.9%; Score 11; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALNI 638
||||||||||||
Db 633 IIPYIGPALNI 643

RESULT 13
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-28

Query Match 0.9%; Score 11; DB 2; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALNI 638
||||||||||||
Db 633 IIPYIGPALNI 643

RESULT 14
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321

```
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             0.9%;  Score 11;  DB 3;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||| 
Db        633 IIPYIGPALNI 643


RESULT 15
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             0.9%;  Score 11;  DB 4;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||| 
Db        633 IIPYIGPALNI 643


Search completed:  September 10, 2004, 05:17:08
Job time :  22.1666 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 65.3048 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 1291
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1291	100.0	1291	12	US-10-272-898-60	Sequence 60, Appl
2	1291	100.0	1291	15	US-10-354-774-60	Sequence 60, Appl
3	1291	100.0	1291	15	US-10-271-012-60	Sequence 60, Appl
4	1291	100.0	1291	15	US-10-452-024-135	Sequence 135, App
5	1291	100.0	1291	15	US-10-205-516-6	Sequence 6, Appli
6	1291	100.0	1291	16	US-10-729-122-60	Sequence 60, Appl
7	1291	100.0	1291	16	US-10-729-039-60	Sequence 60, Appl
8	1206	93.4	1291	15	US-10-452-024-3	Sequence 3, Appli
9	1206	93.4	1291	15	US-10-452-024-136	Sequence 136, App
10	991	76.8	1301	15	US-10-205-516-20	Sequence 20, Appl
11	866	67.1	876	14	US-10-241-596-128	Sequence 128, App
12	866	67.1	877	14	US-10-241-596-126	Sequence 126, App
13	866	67.1	877	14	US-10-241-596-130	Sequence 130, App
14	866	67.1	881	14	US-10-241-596-124	Sequence 124, App
15	866	67.1	902	14	US-10-241-596-132	Sequence 132, App
16	866	67.1	912	14	US-10-241-596-116	Sequence 116, App
17	866	67.1	914	14	US-10-241-596-120	Sequence 120, App
18	866	67.1	944	14	US-10-241-596-122	Sequence 122, App
19	866	67.1	950	14	US-10-241-596-118	Sequence 118, App
20	552	42.8	552	15	US-10-452-024-141	Sequence 141, App
21	449	34.8	450	10	US-09-910-186A-10	Sequence 10, Appl
22	444	34.4	444	13	US-10-011-588-9	Sequence 9, Appli
23	436	33.8	436	13	US-10-011-588-27	Sequence 27, Appl
24	436	33.8	462	12	US-10-272-898-62	Sequence 62, Appl
25	436	33.8	462	15	US-10-354-774-62	Sequence 62, Appl
26	436	33.8	462	15	US-10-271-012-62	Sequence 62, Appl
27	436	33.8	462	16	US-10-729-122-62	Sequence 62, Appl
28	436	33.8	462	16	US-10-729-039-62	Sequence 62, Appl
29	436	33.8	811	13	US-10-011-588-29	Sequence 29, Appl
30	398	30.8	399	10	US-09-910-186A-24	Sequence 24, Appl
31	119	9.2	1280	15	US-10-452-024-139	Sequence 139, App
32	119	9.2	1280	15	US-10-452-024-140	Sequence 140, App
33	68	5.3	386	10	US-09-910-186A-26	Sequence 26, Appl
34	68	5.3	824	13	US-10-011-588-33	Sequence 33, Appl
35	68	5.3	1275	15	US-10-452-024-167	Sequence 167, App
36	68	5.3	1276	12	US-10-272-898-66	Sequence 66, Appl
37	68	5.3	1276	15	US-10-354-774-66	Sequence 66, Appl
38	68	5.3	1276	15	US-10-271-012-66	Sequence 66, Appl
39	68	5.3	1276	15	US-10-452-024-4	Sequence 4, Appli
40	68	5.3	1276	15	US-10-452-024-164	Sequence 164, App
41	68	5.3	1276	15	US-10-452-024-168	Sequence 168, App
42	68	5.3	1276	15	US-10-205-516-8	Sequence 8, Appli
43	68	5.3	1276	16	US-10-729-122-66	Sequence 66, Appl
44	68	5.3	1276	16	US-10-729-039-66	Sequence 66, Appl
45	68	5.3	1285	15	US-10-452-024-165	Sequence 165, App

ALIGNMENTS

RESULT 1
US-10-272-898-60
; Sequence 60, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

```
;           LENGTH: 1291 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-272-898-60

  Query Match            100.0%;  Score 1291;  DB 12;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 2
US-10-354-774-60
; Sequence 60, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-354-774-60

  Query Match            100.0%;  Score 1291;  DB 15;  Length 1291;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 3
US-10-271-012-60
; Sequence 60, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-271-012-60

  Query Match           100.0%;  Score 1291;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
```

```
                                                
```

```
Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 5
US-10-205-516-6
; Sequence 6, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-6

Query Match 100.0%; Score 1291; DB 15; Length 1291;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 6
US-10-729-122-60
; Sequence 60, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-122-60

  Query Match           100.0%;  Score 1291;  DB 16;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
             |||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 7
US-10-729-039-60
; Sequence 60, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-039-60

  Query Match           100.0%;  Score 1291;  DB 16;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||||
```

```
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 8
US-10-452-024-3
; Sequence 3, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-3

  Query Match            93.4%;  Score 1206;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         86 FLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 FLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNN 145

Qy        146 WVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNA 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 WVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNA 205

Qy        206 TNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKL 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 TNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKL 265

Qy        266 EYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQK 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 EYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQK 325

Qy        326 LIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTA 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 LIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTA 385

Qy        386 NILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRS 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 NILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRS 445

Qy        446 LYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKN 505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 LYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKN 505

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565

Qy        566 RSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 RSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625

Qy        626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685

Qy        686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745

Qy        746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805

Qy        806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865

Qy        866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQ 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQ 925

Qy        926 NENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 NENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 985

Qy        986 DSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 DSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1045

Qy       1046 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1105

Qy       1106 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1165

Qy       1166 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTY 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTY 1225

Qy       1226 YYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1226 YYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWG 1285

Qy       1286 FVPVSE 1291
              ||||||
Db       1286 FVPVSE 1291


RESULT 9
US-10-452-024-136
; Sequence 136, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 136
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-136

  Query Match            93.4%;  Score 1206;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         86 FLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 FLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNN 145

Qy        146 WVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNA 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        146 WVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNA 205

Qy        206 TNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKL 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 TNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKL 265

Qy        266 EYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQK 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 EYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQK 325

Qy        326 LIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTA 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 LIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTA 385

Qy        386 NILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRS 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 NILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRS 445

Qy        446 LYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKN 505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 LYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKN 505

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565

Qy        566 RSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 RSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625

Qy        626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685

Qy        686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745

Qy        746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805

Qy        806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865

Qy        866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQ 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQ 925

Qy        926 NENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 NENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 985

Qy        986 DSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 DSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1045

Qy       1046 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1105

Qy       1106 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1165

Qy       1166 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTY 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTY 1225

Qy       1226 YYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1226 YYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWG 1285

Qy       1286 FVPVSE 1291
              ||||||
Db       1286 FVPVSE 1291


RESULT 10
US-10-205-516-20
; Sequence 20, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 1301
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-20

  Query Match            76.8%;  Score 991;  DB 15;  Length 1301;
  Best Local Similarity  99.7%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    4;  Gaps    1;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNK----TLDCRELLVKNTDLPFIGDISDVKTDI 476
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKIEGRTLDCRELLVKNTDLPFIGDISDVKTDI 480

Qy        477 FLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDN 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDN 540

Qy        537 RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGG 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGG 600

Qy        597 LFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGV 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGV 660
```

```
Qy        657 TILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRI 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRI 720

Qy        717 ITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAM 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAM 780

Qy        777 NNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKV 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKV 840

Qy        837 NNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEE 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEE 900

Qy        897 GDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGY 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGY 960

Qy        957 TIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNM 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNM 1020

Qy       1017 MGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIF 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 MGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIF 1080

Qy       1077 AKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTR 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTR 1140

Qy       1137 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 1200

Qy       1197 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 1260

Qy       1257 DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||||||||
Db       1261 DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1295
```

```
RESULT 11
US-10-241-596-128
; Sequence 128, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 128
;   LENGTH: 876
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-128

  Query Match           67.1%;  Score 866;  DB 14;  Length 876;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  866;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 70

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 130

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 190

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 250

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 310

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 370

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 430

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        431 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 490

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        491 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 550

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        551 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 610

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 670

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        671 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 730

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        731 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 790

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        791 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 850

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN 866
              |||||||||||||||||||||||||
Db        851 QNTIPFNIFSYTNNSLLKDIINEYFN 876
```

```
RESULT 12
US-10-241-596-126
; Sequence 126, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
;   LENGTH: 877
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-126

  Query Match             67.1%;  Score 866;  DB 14;  Length 877;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  866;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 71

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 131

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 191

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 251

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 311

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 371

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 431

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 491

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 551

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 611

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 671

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 731

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 791

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 851

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN 866
              |||||||||||||||||||||||||
Db        852 QNTIPFNIFSYTNNSLLKDIINEYFN 877


RESULT 13
US-10-241-596-130
; Sequence 130, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
;   LENGTH: 877
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-130

  Query Match             67.1%;  Score 866;  DB 14;  Length 877;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  866;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 71

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 131

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 191

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 251

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 311
```

```
Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 371

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 431

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 491

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 551

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 611

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 671

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 731

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 791

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 851

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN 866
              |||||||||||||||||||||||||
Db        852 QNTIPFNIFSYTNNSLLKDIINEYFN 877
```

RESULT 14
US-10-241-596-124
; Sequence 124, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 124
; LENGTH: 881
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-124

Query Match 67.1%; Score 866; DB 14; Length 881;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 866; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 75

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 135

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 195

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 255

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 315

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 375

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 435

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 495

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 555

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 615

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 675

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 735

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 795

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 855

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN 866
              |||||||||||||||||||||||||
Db        856 QNTIPFNIFSYTNNSLLKDIINEYFN 881
```

RESULT 15
US-10-241-596-132
; Sequence 132, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003

```
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 132
;   LENGTH: 902
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-132

  Query Match           67.1%;  Score 866;  DB 14;  Length 902;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  866;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 75

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 135

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 195

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 255

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 315

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 375

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 435

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 495

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 555

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 615

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 675

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 735

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 795

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 855

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN 866
              |||||||||||||||||||||||||
Db        856 QNTIPFNIFSYTNNSLLKDIINEYFN 881


Search completed: September 10, 2004, 05:52:46
Job time : 69.3048 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 17.885 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 1291
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1291	100.0	1291	2	S46431	botulinum neurotox
2	1291	100.0	1291	2	A49777	botulinum neurotox
3	552	42.8	552	2	A49928	neurotoxin C1, BoN
4	68	5.3	1276	2	S11455	botulinum neurotox
5	68	5.3	1285	2	S70582	botulinum neurotox
6	24	1.9	24	2	A43503	botulinum toxin ty
7	11	0.9	1296	1	BTCLAB	bontoxilysin (EC 3
8	10	0.8	1297	2	S39791	neurotoxin - Clost
9	9	0.7	512	2	B87804	protein K04F10.2 [
10	9	0.7	548	2	C70253	adenine deaminase
11	9	0.7	603	2	F69006	DNA-directed RNA p
12	9	0.7	604	2	S02195	DNA-directed RNA p
13	9	0.7	1163	2	E71481	probable swf/snf h
14	9	0.7	1274	2	I40813	neurotoxin type F
15	9	0.7	1296	2	I40645	botulinum neurotox
16	9	0.7	1315	1	BTCLTN	tentoxilysin (EC 3
17	8	0.6	110	2	T37490	ribosomal protein
18	8	0.6	255	2	H86321	hypothetical prote
19	8	0.6	255	2	G71977	probable outer mem
20	8	0.6	271	2	T22770	hypothetical prote
21	8	0.6	310	2	A70421	fructose-1,6-bisph
22	8	0.6	334	2	B75095	probable histone d
23	8	0.6	582	2	S57722	sli-1 protein - Ca
24	8	0.6	585	2	T29654	hypothetical prote
25	8	0.6	673	2	AF1143	internalin protein
26	8	0.6	1051	2	A38373	ubiquitin-protein
27	8	0.6	1053	2	T06483	probable ubiquitin
28	8	0.6	1080	2	T03964	probable ubiquitin
29	8	0.6	1080	2	T00587	probable ubiquitin
30	8	0.6	1291	1	A48940	bontoxilysin (EC 3
31	8	0.6	1474	2	F69009	probable membrane
32	7	0.5	37	2	A60117	neurotoxin type D
33	7	0.5	38	2	B60117	neurotoxin type D
34	7	0.5	68	2	G83963	hypothetical prote
35	7	0.5	91	2	AF2068	hypothetical prote
36	7	0.5	95	2	G90945	hypothetical prote
37	7	0.5	95	2	B85794	hypothetical prote
38	7	0.5	95	2	A64944	hypothetical prote
39	7	0.5	98	2	S28243	NADH2 dehydrogenas
40	7	0.5	99	2	H90594	hypothetical prote
41	7	0.5	99	2	F70154	conserved hypothet
42	7	0.5	104	2	T09471	hypothetical prote
43	7	0.5	105	2	AD2181	hypothetical prote
44	7	0.5	106	1	R6DOP2	acidic ribosomal p
45	7	0.5	135	2	G70128	conserved hypothet

ALIGNMENTS

RESULT 1
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 100.0%; Score 1291; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
```

```
Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 2
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.;
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxir
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutinir
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whol
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 100.0%; Score 1291; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
```

```
Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 3
A49928
neurotoxin C1, BoNT/C1 - Clostridium botulinum phage CN
C;Species: Clostridium botulinum phage CN
C;Date: 13-Sep-1994 #sequence_revision 25-Apr-1997 #text_change 28-Jul-2000
C;Accession: A49928
R;Hauser, D.; Gibert, M.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
J. Bacteriol. 175, 7260-7268, 1993
A;Title: Comparative analysis of C3 and botulinal neurotoxin genes and their environment
A;Reference number: A49928; MUID:94042900; PMID:8226673
A;Accession: A49928
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-552 <HAU>
A;Note: sequence extracted from NCBI backbone (NCBIN:139803, NCBIP:139804)
A;Note: strain A028
C;Superfamily: tetanus toxin

```
  Query Match            42.8%;  Score 552;  DB 2;  Length 552;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  552;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLES 552
              ||||||||||||
Db        541 VDYLNSYYYLES 552
```

RESULT 4
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

```
  Query Match             5.3%;  Score 68;  DB 2;  Length 1276;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-62;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 792

Qy        797 PKVIDELN 804
              ||||||||
Db        793 PKVIDELN 800
```

RESULT 5
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Morlishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membr
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

Query Match 5.3%; Score 68; DB 2; Length 1285;
Best Local Similarity 100.0%; Pred. No. 4.3e-62;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 733 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 792

Qy 797 PKVIDELN 804
||||||||
Db 793 PKVIDELN 800

RESULT 6
A43503
botulinum toxin type C light chain - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 21-Oct-1992 #sequence_revision 21-Oct-1992 #text_change 21-Aug-1998
C;Accession: A43503
R;Tsuzuki, K.; Yokosawa, N.; Syuto, B.; Ohishi, I.; Fujii, N.; Kimura, K.; Oguma, K.
Infect. Immun. 56, 898-902, 1988
A;Title: Establishment of a monoclonal antibody recognizing an antigenic site common to
A;Reference number: A43503; MUID:88153072; PMID:2450068
A;Accession: A43503
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-24 <TSU>
C;Superfamily: tetanus toxin

Query Match 1.9%; Score 24; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ITINNFNYSDPVDNKNILYLDTHL 26
||||||||||||||||||||||||
Db 1 ITINNFNYSDPVDNKNILYLDTHL 24

RESULT 7
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.,
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin,
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, anc
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ext
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arour
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; N
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associatec
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prote
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 0.9%; Score 11; DB 1; Length 1296;

Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALNI 638
||||||||||
Db 633 IIPYIGPALNI 643

RESULT 8
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 0.8%; Score 10; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 IIPYIGPALN 637
||||||||||
Db 625 IIPYIGPALN 634

RESULT 9
B87804
protein K04F10.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: B87804
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: B87804
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:chr_I; PIDN:AAB96751.1; PID:g2773240; GSPDB:GN00019; CESP:K04F10.
C;Genetics:
A;Gene: K04F10.2
A;Map position: 1

Query Match 0.7%; Score 9; DB 2; Length 512;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1021 KIYINGKLI 1029
|||||||||
Db 313 KIYINGKLI 321

RESULT 10
C70253
adenine deaminase (adeC) homolog - Lyme disease spirochete plasmid K/lp36
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 29-Sep-1999
C;Accession: C70253
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: C70253
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-548 <KLE>
A;Cross-references: GB:AE000788; NID:g2690123; PIDN:AAC66139.1; PID:g2690131; TIGR:BBK17
A;Experimental source: strain B31
C;Genetics:
A;Genome: plasmid
C;Superfamily: adenine deaminase adeC

Query Match 0.7%; Score 9; DB 2; Length 548;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 819 DSHNIILVG 827
|||||||||
Db 438 DSHNIILVG 446

RESULT 11
F69006
DNA-directed RNA polymerase (EC 2.7.7.6) chain B' - Methanobacterium thermoautotrophicum
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Dec-2000
C;Accession: F69006
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: F69006
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-603 <MTH>
A;Cross-references: GB:AE000876; GB:AE000666; NID:g2622140; PIDN:AAB85541.1; PID:g262214
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1050
A;Start codon: GTG
C;Superfamily: Halobacterium salinarum DNA-directed RNA polymerase chain B'
C;Keywords: nucleotidyltransferase; transcription

Query Match 0.7%; Score 9; DB 2; Length 603;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1021 KIYINGKLI 1029
|||||||||
Db 5 KIYINGKLI 13

RESULT 12
S02195
DNA-directed RNA polymerase (EC 2.7.7.6) chain B' - Methanobacterium thermoautotrophicum
N;Alternate names: transcriptase B' chain
C;Species: Methanobacterium thermoautotrophicum
A;Variety: strain Winter
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 08-Dec-2000
C;Accession: S02195; B28213
R;Berghoefer, B.; Kroeckel, L.; Koertner, C.; Truss, M.; Schallenberg, J.; Klein, A.
Nucleic Acids Res. 16, 8113-8128, 1988
A;Title: Relatedness of archaebacterial RNA polymerase core subunits to their eubacterial
A;Reference number: S02194; MUID:88335550; PMID:2843811
A;Accession: S02195
A;Molecule type: DNA
A;Residues: 1-604 <BER>
A;Cross-references: EMBL:X08038; NID:g44664; PIDN:CAA30837.1; PID:g809726

A;Experimental source: strain Winter
R;Schallenberg, J.; Moes, M.; Truss, M.; Reiser, W.; Thomm, M.; Stetter, K.O.; Klein, A.
J. Bacteriol. 170, 2247-2253, 1988
A;Title: Cloning and physical mapping of RNA polymerase genes from Methanobacterium ther
anogenic archaebacteria.
A;Reference number: A28213; MUID:88198019; PMID:2834336
A;Accession: B28213
A;Molecule type: DNA
A;Residues: 'FFPPGGCYV',2-55 <SCH>
A;Cross-references: GB:M20391; NID:g149737; PIDN:AAA72653.1; PID:g551893
A;Experimental source: strain Winter
A;Note: the sequence published in Fig. 2 is shown; the GenBank entry MBFRNAPB (PID:g5518
C;Genetics:
A;Start codon: GTG
C;Superfamily: Halobacterium salinarum DNA-directed RNA polymerase chain B'
C;Keywords: nucleotidyltransferase; transcription

  Query Match 0.7%; Score 9; DB 2; Length 604;
  Best Local Similarity 100.0%; Pred. No. 2.2;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1021 KIYINGKLI 1029
|||||||||
Db 5 KIYINGKLI 13

RESULT 13
E71481
probable swf/snf helicase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: E71481
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: E71481
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1163 <ARN>
A;Cross-references: GB:AE001341; GB:AE001273; NID:g3329156; PIDN:AAC68303.1; PID:g332916
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT708

  Query Match 0.7%; Score 9; DB 2; Length 1163;
  Best Local Similarity 100.0%; Pred. No. 3.9;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 872 KILSLQNRK 880
|||||||||
Db 1126 KILSLQNRK 1134

RESULT 14
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 0.7%; Score 9; DB 2; Length 1274;
  Best Local Similarity 100.0%; Pred. No. 4.3;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 IKLFKRINS 99
|||||||||
Db 92 IKLFKRINS 100

RESULT 15
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A a
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 0.7%; Score 9; DB 2; Length 1296;
  Best Local Similarity 100.0%; Pred. No. 4.3;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 630 PYIGPALNI 638
|||||||||
Db 635 PYIGPALNI 643

Search completed: September 10, 2004, 05:14:52
Job time : 20.985 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 11.1576 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 1291
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1206	93.4	1290	1	BXC1_CLOBO	P18640 clostridium
2	68	5.3	1276	1	BXD_CLOBO	P19321 clostridium
3	11	0.9	1295	1	BXA1_CLOBO	P10845 clostridium
4	10	0.8	1296	1	BXG_CLOBO	Q60393 clostridium
5	9	0.7	548	1	ADEC_BORBU	O50821 borrelia bu
6	9	0.7	603	1	RPB1_METTH	O27124 methanobact
7	9	0.7	604	1	RPB1_METTW	P09845 methanobact
8	9	0.7	1274	1	BXF_CLOBO	P30996 clostridium
9	9	0.7	1295	1	BXA2_CLOBO	Q45894 clostridium
10	9	0.7	1314	1	TETX_CLOTE	P04958 clostridium
11	8	0.6	621	1	TP6B_METAC	Q8tqf7 methanosarc
12	8	0.6	1051	1	UBA1_WHEAT	P20973 triticum ae
13	8	0.6	1051	1	UBA2_WHEAT	P31251 triticum ae
14	8	0.6	1053	1	UBA3_WHEAT	P31252 triticum ae
15	8	0.6	1290	1	BXB_CLOBO	P10844 clostridium
16	7	0.5	68	1	RPOZ_BACHD	Q9k9y3 bacillus ha
17	7	0.5	95	1	YEBO_ECO57	Q8xcn8 escherichia
18	7	0.5	95	1	YEBO_ECOLI	P76266 escherichia
19	7	0.5	97	1	NB2M_BOVIN	Q02365 bos taurus
20	7	0.5	99	1	Y439_BORBU	P33762 borrelia bu
21	7	0.5	105	1	RLA2_DICDI	P22683 dictyosteli
22	7	0.5	118	1	RS13_WIGBR	Q8d1z0 wiggleswort
23	7	0.5	143	1	ATP9_EMENI	P16000 emericella
24	7	0.5	146	1	PHEB_THETN	Q8r708 thermoanaer
25	7	0.5	164	1	COAD_METTH	O27918 methanobact
26	7	0.5	165	1	Y084_RICPR	Q9ze63 rickettsia
27	7	0.5	168	1	VAT_CERV	P05397 carnation e
28	7	0.5	178	1	ATPD_BACCA	P41011 bacillus ca
29	7	0.5	179	1	ATPD_BACP3	P09220 bacillus ps
30	7	0.5	188	1	YP34_DEIRA	Q9rrf8 deinococcus
31	7	0.5	204	1	CLPP_NEIMA	Q9ju33 neisseria m
32	7	0.5	204	1	CLPP_NEIMB	Q9jz38 neisseria m
33	7	0.5	216	1	Y458_METJA	Q57900 methanococc
34	7	0.5	233	1	RSUA_VIBVU	Q8d8x2 vibrio vuln
35	7	0.5	240	1	THIE_METMA	Q8ps49 methanosarc
36	7	0.5	255	1	DAPB_CLOTE	Q891s2 clostridium
37	7	0.5	257	1	CODY_STAAM	Q99ul6 staphylococ
38	7	0.5	265	1	YFU2_BACST	Q04729 bacillus st
39	7	0.5	270	1	RECX_BACCR	Q81i97 bacillus ce
40	7	0.5	275	1	PORI_DICDI	Q01501 dictyosteli
41	7	0.5	278	1	CHIS_NOCSP	P48846 nocardioide
42	7	0.5	290	1	YOAU_BACSU	O34701 bacillus su
43	7	0.5	293	1	Y844_ARCFU	O29414 archaeoglob
44	7	0.5	307	1	ENGC_BORBU	O51126 borrelia bu
45	7	0.5	308	1	PAP2_FOWPV	P15916 fowlpox vir

ALIGNMENTS

RESULT 1
BXC1_CLOBO
ID BXC1_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE (Bontoxilysin C1).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90370487; PubMed=2204031;
RA Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA Boquet P., Popoff M.R.;
RT "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL Nucleic Acids Res. 18:4924-4924(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=91024998; PubMed=2222445;
RA Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT "The complete nucleotide sequence of the gene coding for botulinum
RT type C1 toxin in the C-ST phage genome.";
RL Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN [3]
RP SEQUENCE OF 2-25.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=88153072; PubMed=2450068;
RA Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA Oguma K.;
RT "Establishment of a monoclonal antibody recognizing an antigenic site
RT common to Clostridium botulinum type B, C1, D, and E toxins and
RT tetanus toxin.";
RL Infect. Immun. 56:898-902(1988).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94038966; PubMed=7901002;
RA Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT cleaving HPC-1/syntaxin.";
RL EMBO J. 12:4821-4828(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             93.4%;  Score 1206;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         86 FLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 FLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNN 144

Qy        146 WVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNA 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 WVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNA 204

Qy        206 TNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKL 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 TNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKL 264

Qy        266 EYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQK 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 EYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQK 324

Qy        326 LIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTA 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 LIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTA 384

Qy        386 NILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRS 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 NILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRS 444

Qy        446 LYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKN 505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 LYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKN 504

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 564

Qy        566 RSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 RSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 624

Qy        626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 684

Qy        686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 744

Qy        746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 804

Qy        806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 864

Qy        866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQ 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQ 924

Qy        926 NENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 NENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 984

Qy        986 DSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 DSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1044

Qy       1046 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1104

Qy       1106 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1164

Qy       1166 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTY 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTY 1224

Qy       1226 YYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 YYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWG 1284

Qy       1286 FVPVSE 1291
              ||||||
Db       1285 FVPVSE 1290


RESULT 2
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
```

```
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             5.3%;  Score 68;  DB 1;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-62;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 792

Qy        797 PKVIDELN 804
              ||||||||
Db        793 PKVIDELN 800


RESULT 3
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
```

```
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
```

```
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             0.9%;  Score 11;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.016;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALNI 638
              ||||||||||| 
Db        632 IIPYIGPALNI 642


RESULT 4
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             0.8%;  Score 10;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 IIPYIGPALN 637
              ||||||||||
Db        624 IIPYIGPALN 633


RESULT 5
ADEC_BORBU
ID   ADEC_BORBU     STANDARD;      PRT;   548 AA.
AC   O50821;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable adenine deaminase (EC 3.5.4.2) (Adenase) (Adenine aminase).
GN   BBK17.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid lp36.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- CATALYTIC ACTIVITY: Adenine + H(2)O = hypoxanthine + NH(3).
CC   -!- SIMILARITY: Belongs to the adenine deaminase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000788; AAC66139.1; -.
DR   PIR; C70253; C70253.
DR   TIGR; BBK17; -.
DR   InterPro; IPR006679; Adenine_deam.
DR   InterPro; IPR006680; Amidohydro_1.
DR   InterPro; IPR005847; Pept_M38_regn.
DR   Pfam; PF01979; Amidohydro_1; 1.
DR   ProDom; PD000518; Urease; 1.
DR   TIGRFAMs; TIGR01178; ade; 1.
KW   Hypothetical protein; Hydrolase; Plasmid; Complete proteome.
SQ   SEQUENCE   548 AA;  60864 MW;  012AE5222AD8301E CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 0.96;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        819 DSHNIILVG 827
              |||||||||
Db        438 DSHNIILVG 446


RESULT 6
RPB1_METTH
ID   RPB1_METTH     STANDARD;      PRT;   603 AA.
AC   O27124;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA-directed RNA polymerase subunit B' (EC 2.7.7.6).
GN   RPOB1 OR MTH1050.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=187420;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Delta H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES. THE B (B''+B' AND BETA) SUBUNITS HAVE BEEN IMPLICATED
CC       IN DNA PROMOTER RECOGNITION AND ALSO IN NUCLEOTIDE BINDING.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- COFACTOR: Zinc.
CC   -!- SUBUNIT: M.thermoautotrophicum RNA polymerase is composed of four
CC       subunits: A',A'',B' AND B''.
CC   -!- SIMILARITY: THE COMBINED B"+B' SUBUNITS ARE EQUIVALENT TO THE B
CC       SUBUNITS OF THE EUKARYOTIC POLYMERASES I AND II AND TO EUBACTERIAL
CC       BETA SUBUNIT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000876; AAB85541.1; -.
DR   PIR; F69006; F69006.
DR   HSSP; Q9KWU7; 1HQM.
DR   InterPro; IPR007121; RNA_pol_B.
DR   InterPro; IPR007646; RNA_pol_Rpb2_4.
DR   InterPro; IPR007647; RNA_pol_Rpb2_5.
DR   InterPro; IPR007120; RNA_pol_Rpb2_6.
DR   InterPro; IPR007641; RNA_pol_Rpb2_7.
DR   Pfam; PF04566; RNA_pol_Rpb2_4; 1.
DR   Pfam; PF04567; RNA_pol_Rpb2_5; 1.
DR   Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR   Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase; Zinc;
KW   Complete proteome; Zinc-finger.
FT   ZN_FING     546    567       POTENTIAL.
SQ   SEQUENCE   603 AA;  67687 MW;  17F2C0FC70BD5BBC CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 603;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1021 KIYINGKLI 1029
              |||||||||
Db          5 KIYINGKLI 13


RESULT 7
RPB1_METTW
ID   RPB1_METTW     STANDARD;      PRT;   604 AA.
AC   P09845; P72014;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA-directed RNA polymerase subunit B' (EC 2.7.7.6).
GN   RPOB1 OR RPOU.
OS   Methanobacterium thermoautotrophicum (strain Winter).
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=79930;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88335550; PubMed=2843811;
RA   Berghoefer B., Kroeckel L., Koertner C., Truss M., Schallenberg J.,
RA   Klein A.;
RT   "Relatedness of archaebacterial RNA polymerase core subunits to their
RT   eubacterial and eukaryotic equivalents.";
RL   Nucleic Acids Res. 16:8113-8128(1988).
RN   [2]
RP   SEQUENCE OF 1-55 FROM N.A.
RX   MEDLINE=88198019; PubMed=2834336;
RA   Schallenberg J., Moes M., Truss M., Reiser W., Thomm M.,
RA   Stetter K.O., Klein A.;
RT   "Cloning and physical mapping of RNA polymerase genes from
RT   Methanobacterium thermoautotrophicum and comparison of homologies and
RT   gene orders with those of RNA polymerase genes from other
RT   methanogenic archaebacteria.";
RL   J. Bacteriol. 170:2247-2253(1988).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES. THE B (B''+B' AND BETA) SUBUNITS HAVE BEEN IMPLICATED
CC       IN DNA PROMOTER RECOGNITION AND ALSO IN NUCLEOTIDE BINDING.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- COFACTOR: Zinc.
CC   -!- SUBUNIT: M.thermoautotrophicum RNA polymerase is composed of four
CC       subunits: A',A'',B' AND B''.
CC   -!- SIMILARITY: THE COMBINED B"+B' SUBUNITS ARE EQUIVALENT TO THE B
CC       SUBUNITS OF THE EUKARYOTIC POLYMERASES I AND II AND TO EUBACTERIAL
CC       BETA SUBUNIT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X08038; CAA30837.1; -.
DR   EMBL; M20391; AAA72653.1; -.
DR   PIR; S02195; S02195.
DR   HSSP; Q9KWU7; 1HQM.
DR   InterPro; IPR007121; RNA_pol_B.
DR   InterPro; IPR007646; RNA_pol_Rpb2_4.
DR   InterPro; IPR007647; RNA_pol_Rpb2_5.
DR   InterPro; IPR007120; RNA_pol_Rpb2_6.
DR   InterPro; IPR007641; RNA_pol_Rpb2_7.
DR   Pfam; PF04566; RNA_pol_Rpb2_4; 1.
DR   Pfam; PF04567; RNA_pol_Rpb2_5; 1.
DR   Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR   Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase; Zinc;
KW   Zinc-finger.
FT   ACT_SITE    348    348       BY SIMILARITY.
FT   ACT_SITE    481    481       BY SIMILARITY.
FT   ZN_FING     547    568       POTENTIAL.
SQ   SEQUENCE   604 AA;  68113 MW;  97177D1F9E15A23D CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 604;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1021 KIYINGKLI 1029
              |||||||||
Db          5 KIYINGKLI 13


RESULT 8
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 IKLFKRINS 99
              |||||||||
Db         92 IKLFKRINS 100
```

```
RESULT 9
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             0.7%;  Score 9;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        630 PYIGPALNI 638
              |||||||||
Db        634 PYIGPALNI 642


RESULT 10
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
```

```
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
```

```
FT   STRAND      1148   1152
FT   STRAND      1155   1158
FT   TURN        1159   1162
FT   STRAND      1163   1166
FT   STRAND      1173   1178
FT   TURN        1184   1185
FT   STRAND      1188   1188
FT   STRAND      1190   1190

  Query Match             0.7%;  Score 9;  DB 1;  Length 1314;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        630 PYIGPALNI 638
              |||||||||
Db        643 PYIGPALNI 651


RESULT 11
TP6B_METAC
ID   TP6B_METAC     STANDARD;      PRT;   621 AA.
AC   Q8TQF7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Type II DNA topoisomerase VI subunit B (EC 5.99.1.3) (TopoVI-B).
GN   TOP6B OR MA1587.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
CC   -!- FUNCTION: Relaxes both positive and negative superturns and
CC       exhibits a strong decatenase activity. The B subunit binds ATP (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC       of double-stranded DNA.
CC   -!- SUBUNIT: Heterotetramer of two subunits A and two subunits B (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the TOP6B family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010829; AAM05000.1; -.
DR   HAMAP; MF_00322; -; 1.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR005734; DNA_top6B.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   TIGRFAMs; TIGR01052; top6b; 1.
KW   Isomerase; Topoisomerase; DNA-binding; ATP-binding; Complete proteome.
SQ   SEQUENCE   621 AA;  68908 MW;  3344AF34A1BD0044 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 621;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        102 IGEELIYR 109
              ||||||||
Db        320 IGEELIYR 327


RESULT 12
UBA1_WHEAT
ID   UBA1_WHEAT     STANDARD;      PRT;  1051 AA.
AC   P20973;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ubiquitin-activating enzyme E1 1.
GN   UBA1.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=cv. Augusta;
RX   MEDLINE=90368797; PubMed=2203788;
RA   Hatfield P.M., Callis J., Vierstra R.D.;
RT   "Cloning of ubiquitin activating enzyme from wheat and expression of
RT   a functional protein in Escherichia coli.";
RL   J. Biol. Chem. 265:15813-15817(1990).
RN   [2]
RP   MUTAGENESIS OF CYSTEINE RESIDUES, AND ACTIVE SITE.
RX   MEDLINE=92340519; PubMed=1634524;
RA   Hatfield P.M., Vierstra R.D.;
RT   "Multiple forms of ubiquitin-activating enzyme E1 from wheat.
RT   Identification of an essential cysteine by in vitro mutagenesis.";
RL   J. Biol. Chem. 267:14799-14803(1992).
CC   -!- FUNCTION: Activates ubiquitin by first adenylating with ATP its
CC       carboxy-terminal glycine residue and thereafter linking this
CC       residue to the side chain of a cysteine residue in E1, yielding an
CC       ubiquitin-E1 thiolester and free AMP.
CC   -!- PATHWAY: Ubiquitin conjugation; first step.
CC   -!- SUBUNIT: Monomer.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- MISCELLANEOUS: There are two active sites within the E1 molecule,
CC       allowing it to accomodate two ubiquitin moieties at a time, with a
CC       new ubiquitin forming an adenylate intermediate as the previous
CC       one is transferred to the thiol site.
CC   -!- MISCELLANEOUS: There are multiple genes encoding E1 in wheat.
CC   -!- SIMILARITY: Belongs to the ubiquitin-activating E1 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M55604; AAA34308.1; -.
DR   InterPro; IPR009036; MoeB.
DR   InterPro; IPR000594; ThiF_domain.
DR   InterPro; IPR000127; UBact_repeat.
DR   InterPro; IPR000011; Uqtin-activ_enz.
DR   Pfam; PF00899; ThiF; 2.
DR   Pfam; PF02134; UBACT; 2.
DR   TIGRFAMs; TIGR01408; Ube1; 1.
DR   PROSITE; PS00536; UBIQUITIN_ACTIVAT_1; 1.
DR   PROSITE; PS00865; UBIQUITIN_ACTIVAT_2; 1.
```

```
KW   Ubl conjugation pathway; Ligase; Multigene family; Repeat.
FT   ACT_SITE    626    626
FT   DOMAIN       56    605       2 APPROXIMATE REPEATS.
FT   REPEAT       56    194       1-1.
FT   REPEAT      453    605       1-2.
SQ   SEQUENCE   1051 AA;  117007 MW;  9BAB85FE3BACA621 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 1051;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        828 EVDKLKAK 835
              ||||||||
Db        879 EVDKLKAK 886


RESULT 13
UBA2_WHEAT
ID   UBA2_WHEAT     STANDARD;      PRT;  1051 AA.
AC   P31251;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ubiquitin-activating enzyme E1 2.
GN   UBA2.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340519; PubMed=1634524;
RA   Hatfield P.M., Vierstra R.D.;
RT   "Multiple forms of ubiquitin-activating enzyme E1 from wheat.
RT   Identification of an essential cysteine by in vitro mutagenesis.";
RL   J. Biol. Chem. 267:14799-14803(1992).
CC   -!- FUNCTION: Activates ubiquitin by first adenylating with ATP its
CC       carboxy-terminal glycine residue and thereafter linking this
CC       residue to the side chain of a cysteine residue in E1, yielding an
CC       ubiquitin-E1 thiolester and free AMP.
CC   -!- PATHWAY: Ubiquitin conjugation; first step.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- MISCELLANEOUS: There are two active sites within the E1 molecule,
CC       allowing it to accomodate two ubiquitin moieties at a time, with a
CC       new ubiquitin forming an adenylate intermediate as the previous
CC       one is transferred to the thiol site.
CC   -!- MISCELLANEOUS: There are multiple genes encoding E1 in wheat.
CC   -!- SIMILARITY: Belongs to the ubiquitin-activating E1 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M90663; AAA34265.1; -.
DR   InterPro; IPR009036; MoeB.
DR   InterPro; IPR000594; ThiF_domain.
DR   InterPro; IPR000127; UBact_repeat.
DR   InterPro; IPR000011; Uqtin-activ_enz.
DR   Pfam; PF00899; ThiF; 2.
DR   Pfam; PF02134; UBACT; 2.
DR   TIGRFAMs; TIGR01408; Ube1; 1.
DR   PROSITE; PS00536; UBIQUITIN_ACTIVAT_1; 1.
DR   PROSITE; PS00865; UBIQUITIN_ACTIVAT_2; 1.
KW   Ubl conjugation pathway; Ligase; Multigene family; Repeat.
FT   ACT_SITE    626    626       BY SIMILARITY.
FT   DOMAIN       56    605       2 APPROXIMATE REPEATS.
FT   REPEAT       56    194       1-1.
FT   REPEAT      453    605       1-2.
SQ   SEQUENCE   1051 AA;  116825 MW;  B1F8370F6A299B73 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 1051;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        828 EVDKLKAK 835
              ||||||||
Db        879 EVDKLKAK 886


RESULT 14
UBA3_WHEAT
ID   UBA3_WHEAT     STANDARD;      PRT;  1053 AA.
AC   P31252;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ubiquitin-activating enzyme E1 3.
GN   UBA3.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340519; PubMed=1634524;
RA   Hatfield P.M., Vierstra R.D.;
RT   "Multiple forms of ubiquitin-activating enzyme E1 from wheat.
RT   Identification of an essential cysteine by in vitro mutagenesis.";
RL   J. Biol. Chem. 267:14799-14803(1992).
CC   -!- FUNCTION: Activates ubiquitin by first adenylating with ATP its
CC       carboxy-terminal glycine residue and thereafter linking this
CC       residue to the side chain of a cysteine residue in E1, yielding an
CC       ubiquitin-E1 thiolester and free AMP.
CC   -!- PATHWAY: Ubiquitin conjugation; first step.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- MISCELLANEOUS: There are two active sites within the E1 molecule,
CC       allowing it to accomodate two ubiquitin moieties at a time, with a
CC       new ubiquitin forming an adenylate intermediate as the previous
CC       one is transferred to the thiol site.
CC   -!- MISCELLANEOUS: There are multiple genes encoding E1 in wheat.
CC   -!- SIMILARITY: Belongs to the ubiquitin-activating E1 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M90664; AAA34266.1; -.
DR   PIR; T06483; T06483.
DR   InterPro; IPR009036; MoeB.
DR   InterPro; IPR000205; NAD_BS.
DR   InterPro; IPR000594; ThiF_domain.
DR   InterPro; IPR000127; UBact_repeat.
DR   InterPro; IPR000011; Uqtin-activ_enz.
DR   Pfam; PF00899; ThiF; 2.
DR   Pfam; PF02134; UBACT; 2.
DR   TIGRFAMs; TIGR01408; Ube1; 1.
DR   PROSITE; PS00536; UBIQUITIN_ACTIVAT_1; 1.
DR   PROSITE; PS00865; UBIQUITIN_ACTIVAT_2; 1.
KW   Ubl conjugation pathway; Ligase; Multigene family; Repeat.
FT   ACT_SITE    629    629       BY SIMILARITY.
FT   DOMAIN       59    608       2 APPROXIMATE REPEATS.
FT   REPEAT       59    197       1-1.
FT   REPEAT      456    608       1-2.
SQ   SEQUENCE   1053 AA;  116481 MW;  99ABA3F0E85AF7EC CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 1053;
```

```
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        828 EVDKLKAK 835
              ||||||||
Db        881 EVDKLKAK 888


RESULT 15
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1021 KIYINGKL 1028
              ||||||||
```

Db 1014 KIYINGKL 1021

Search completed: September 10, 2004, 05:07:20
Job time : 13.1576 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 52.5064 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 1291
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1291	100.0	1291	2	Q93HT3	Q93ht3 clostridium
2	119	9.2	1280	2	Q45849	Q45849 clostridium
3	119	9.2	1280	2	Q9LBS7	Q9lbs7 clostridium
4	102	7.9	1280	2	Q841S3	Q841s3 clostridium
5	68	5.3	1275	12	Q9QTG7	Q9qtg7 clostridium
6	68	5.3	1285	2	Q45967	Q45967 clostridium
7	68	5.3	1285	2	Q9LBR1	Q9lbr1 clostridium
8	9	0.7	347	2	Q840G4	Q840g4 clostridium
9	9	0.7	568	5	O44770	O44770 caenorhabdi
10	9	0.7	1163	16	O84714	O84714 chlamydia t
11	9	0.7	1166	16	Q821M2	Q821m2 chlamydophi
12	9	0.7	1181	16	Q9PLL8	Q9pll8 chlamydia m
13	9	0.7	1278	2	Q57236	Q57236 clostridium
14	9	0.7	1310	2	Q93N27	Q93n27 clostridium
15	9	0.7	1887	12	Q806Y9	Q806y9 cucumber ye
16	8	0.6	110	3	O14317	O14317 schizosacch
17	8	0.6	122	5	Q86L19	Q86l19 dictyosteli
18	8	0.6	173	10	Q9FSZ0	Q9fsz0 cicer ariet
19	8	0.6	226	12	Q9JG38	Q9jg38 hepatitis b
20	8	0.6	228	10	Q8GXV1	Q8gxv1 arabidopsis
21	8	0.6	237	10	Q7XXK9	Q7xxk9 oryza sativ
22	8	0.6	255	10	Q9M9V0	Q9m9v0 arabidopsis
23	8	0.6	255	16	Q9ZMY9	Q9zmy9 helicobacte
24	8	0.6	256	10	Q8LC68	Q8lc68 arabidopsis
25	8	0.6	271	5	Q20859	Q20859 caenorhabdi
26	8	0.6	289	5	Q8IG38	Q8ig38 caenorhabdi
27	8	0.6	294	16	Q88YK9	Q88yk9 lactobacill
28	8	0.6	307	5	Q94431	Q94431 ciona intes
29	8	0.6	310	16	O67396	O67396 aquifex aeo
30	8	0.6	334	17	Q9UZJ4	Q9uzj4 pyrococcus
31	8	0.6	380	8	Q9T3F5	Q9t3f5 brachyuromy
32	8	0.6	441	2	Q9X708	Q9x708 clostridium
33	8	0.6	493	17	Q8TJF9	Q8tjf9 methanosarc
34	8	0.6	565	5	Q95ZR5	Q95zr5 caenorhabdi
35	8	0.6	582	5	Q19019	Q19019 caenorhabdi
36	8	0.6	655	16	Q87UR4	Q87ur4 pseudomonas
37	8	0.6	673	16	Q8Y9H9	Q8y9h9 listeria mo
38	8	0.6	1045	10	Q84NQ2	Q84nq2 oryza sativ
39	8	0.6	1077	10	P92974	P92974 arabidopsis
40	8	0.6	1080	10	P93028	P93028 arabidopsis
41	8	0.6	1080	10	P93393	P93393 nicotiana t
42	8	0.6	1080	10	Q8RX82	Q8rx82 arabidopsis
43	8	0.6	1161	10	Q9LRZ0	Q9lrz0 arabidopsis
44	8	0.6	1186	16	Q895M7	Q895m7 clostridium
45	8	0.6	1291	2	Q9ZAJ8	Q9zaj8 clostridium

ALIGNMENTS

RESULT 1
Q93HT3
ID Q93HT3 PRELIMINARY; PRT; 1291 AA.
AC Q93HT3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-Yoichi;
RX MEDLINE=21534265; PubMed=11676492;
RA Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,
RA Fujinaga Y., Oguma K., Ohyama T.;
RT "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in
RT Hemagglutination.";
RL Biochem. Biophys. Res. Commun. 288:650-657(2001).
DR EMBL; AB061780; BAB71749.1; -.
DR PIR; A43503; A43503.
DR PIR; S11291; A49777.
DR PIR; S46431; S46431.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  148869 MW;  4A21DB35B8743CF8 CRC64;

  Query Match           100.0%;  Score 1291;  DB 2;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 2
Q45849
ID   Q45849      PRELIMINARY;      PRT;  1280 AA.
AC   Q45849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum C.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=36828;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6813;
RX   MEDLINE=96156810; PubMed=8593068;
RA   Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA   Oguma K.;
RT   "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT   composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT   detection of this gene from Clostridium botulinum type C organisms.";
RL   Appl. Environ. Microbiol. 62:662-667(1996).
DR   EMBL; D49440; BAA08418.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147815 MW;  74F813B228B8C989 CRC64;

  Query Match             9.2%;  Score 119;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-114;
  Matches  119;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        353 ELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMG 412
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 ELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMG 412
```

```
Qy        413 QNLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISD 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 QNLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISD 471


RESULT 3
Q9LBS7
ID   Q9LBS7      PRELIMINARY;      PRT;  1280 AA.
AC   Q9LBS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-6814;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Kawabe T.,
RA   Murakami F., Nakatsuka M., Ohyama T.;
RT   "Organization of Gene Encoding Components of the Botulinum Progenitor
RT   Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT   Chimeric Sequence in the Gene Encoding Each Component.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037166; BAA89713.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147757 MW;  DE124FFB6F68450B CRC64;

  Query Match             9.2%;  Score 119;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-114;
  Matches  119;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        353 ELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMG 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 ELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMG 412

Qy        413 QNLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISD 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 QNLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISD 471


RESULT 4
Q841S3
ID   Q841S3      PRELIMINARY;      PRT;  1280 AA.
AC   Q841S3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2003;
RA   Lee M.S., Cheng M.C., Chen Y.P., Shiau J.R., Lin S.Y.;
RT   "An outbreak of botulism in black-faced spoonbill in Taiwan.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY251553; AAP06952.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1280 AA;  147642 MW;  34C1A2E573EE0AE7 CRC64;

  Query Match             7.9%;  Score 102;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-97;
  Matches  102;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        370 QNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENM 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 QNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNPENM 429

Qy        430 LYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISD 471
              |||||||||||||||||||||||||||||||||||||||||||
Db        430 LYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISD 471


RESULT 5
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
```

```
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match             5.3%;  Score 68;  DB 12;  Length 1275;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-61;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 792

Qy        797 PKVIDELN 804
              ||||||||
Db        793 PKVIDELN 800


RESULT 6
Q45967
ID   Q45967      PRELIMINARY;      PRT;  1285 AA.
AC   Q45967;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE D;
RX   MEDLINE=96283801; PubMed=8679691;
RA   Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA   Ogumad K.;
RT   "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT   types C and D organisms.";
RL   Biochim. Biophys. Acta 1307:123-126(1996).
DR   EMBL; D38442; BAA07477.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1285 AA;  147365 MW;  EEDE98E4ECA6C413 CRC64;

  Query Match             5.3%;  Score 68;  DB 2;  Length 1285;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-61;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 792

Qy        797 PKVIDELN 804
              ||||||||
Db        793 PKVIDELN 800


RESULT 7
Q9LBR1
ID   Q9LBR1      PRELIMINARY;      PRT;  1285 AA.
AC   Q9LBR1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-4947;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA   Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT   "Characterization of the Progenitor Toxin Components Produced by
RT   Clostridium botulinum Type D Strain 4947.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037920; BAA90661.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1285 AA;  147352 MW;  B63AFA487D570680 CRC64;

  Query Match             5.3%;  Score 68;  DB 2;  Length 1285;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-61;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        737 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 AIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNML 792

Qy        797 PKVIDELN 804
              ||||||||
Db        793 PKVIDELN 800


RESULT 8
Q840G4
ID   Q840G4      PRELIMINARY;      PRT;   347 AA.
AC   Q840G4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Toxin (Fragment).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95183076; PubMed=7877632;
RA   Ferreira J.L., Hamdy M.K., McCay S.G., Hemphill M., Kirma N.,
RA   Baumstark B.R.;
RT   "Detection of Clostridium botulinum type F using the polymerase chain
RT   reaction.";
RL   Mol. Cell. Probes 8:365-373(1994).
DR   EMBL; S76749; AAP31981.1; -.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
```

```
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER     347    347
SQ   SEQUENCE   347 AA;  39439 MW;  79AC9C9BF94F4430 CRC64;

  Query Match             0.7%;  Score 9;  DB 2;  Length 347;
  Best Local Similarity   100.0%;  Pred. No. 5.3;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 IKLFKRINS 99
              |||||||||
Db         61 IKLFKRINS 69


RESULT 9
O44770
ID   O44770      PRELIMINARY;      PRT;   568 AA.
AC   O44770;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein K04F10.2.
GN   K04F10.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Wilson R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Latreille P., Wamsley P.;
RT   "The sequence of C. elegans cosmid K04F10.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilson R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF039719; AAB96751.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   568 AA;  64099 MW;  C5A07E185397EA7C CRC64;

  Query Match             0.7%;  Score 9;  DB 5;  Length 568;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1021 KIYINGKLI 1029
              |||||||||
Db        369 KIYINGKLI 377


RESULT 10
O84714
ID   O84714      PRELIMINARY;      PRT;  1163 AA.
AC   O84714;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SWF/SNF family helicase.
GN   CT708.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/Cx;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR   EMBL; AE001341; AAC68303.1; -.
DR   PIR; E71481; E71481.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   InterPro; IPR007527; Znf_SWIM.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   Pfam; PF04434; SWIM; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Complete proteome.
SQ   SEQUENCE   1163 AA;  133178 MW;  42391E5EA27966EE CRC64;

  Query Match             0.7%;  Score 9;  DB 16;  Length 1163;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        872 KILSLQNRK 880
              |||||||||
Db       1126 KILSLQNRK 1134


RESULT 11
Q821M2
ID   Q821M2      PRELIMINARY;      PRT;  1166 AA.
AC   Q821M2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Helicase, Snf2 family.
GN   HELZ OR CCA00918.
OS   Chlamydophila caviae.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83557;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GPIC;
RX   MEDLINE=22569155; PubMed=12682364;
RA   Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA   Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA   Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA   Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA   Fraser C.M.;
RT   "Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC):
RT   examining the role of niche-specific genes in the evolution of the
RT   Chlamydiaceae.";
RL   Nucleic Acids Res. 31:2134-2147(2003).
```

```
DR   EMBL; AE016997; AAP05657.1; -.
DR   TIGR; CCA00918; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   InterPro; IPR007527; Znf_SWIM.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   Pfam; PF04434; SWIM; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Helicase; Complete proteome.
SQ   SEQUENCE   1166 AA;  133629 MW;  373B3C7275D0B243 CRC64;

  Query Match             0.7%;  Score 9;  DB 16;  Length 1166;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        872 KILSLQNRK 880
              |||||||||
Db       1129 KILSLQNRK 1137


RESULT 12
Q9PLL8
ID   Q9PLL8      PRELIMINARY;      PRT;  1181 AA.
AC   Q9PLL8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Helicase, Snf2 family.
GN   TC0081.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn / Nigg;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR   EMBL; AE002276; AAF73530.1; -.
DR   TIGR; TC0081; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   InterPro; IPR007527; Znf_SWIM.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   Pfam; PF04434; SWIM; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Complete proteome.
SQ   SEQUENCE   1181 AA;  135164 MW;  C243F84720F65F88 CRC64;

  Query Match             0.7%;  Score 9;  DB 16;  Length 1181;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        872 KILSLQNRK 880
              |||||||||
Db       1144 KILSLQNRK 1152


RESULT 13
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match             0.7%;  Score 9;  DB 2;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 IKLFKRINS 99
              |||||||||
```

```
Db         92 IKLFKRINS 100


RESULT 14
Q93N27
ID   Q93N27      PRELIMINARY;      PRT;  1310 AA.
AC   Q93N27;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tetanus toxin (Fragment).
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shumin Z., Dianliang L.;
RT   "Cloning and sequence analysis of tetanus toxin gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF389424; AAK72964.2; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER    1310   1310
SQ   SEQUENCE   1310 AA;  150316 MW;  9EADDC914418E450 CRC64;

  Query Match             0.7%;  Score 9;  DB 2;  Length 1310;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        630 PYIGPALNI 638
              |||||||||
Db        645 PYIGPALNI 653


RESULT 15
Q806Y9
ID   Q806Y9      PRELIMINARY;      PRT;  1887 AA.
AC   Q806Y9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Methyltransferase and helicase polyprotein.
OS   Cucumber yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Crinivirus.
OX   NCBI_TaxID=32618;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22541874; PubMed=12655104;
RA   Hartono S., Natsuaki T., Genda Y., Okuda S.;
RT   "Nucleotide sequence and genome organization of Cucumber yellows
RT   virus, a member of the genus Crinivirus.";
RL   J. Gen. Virol. 84:1007-1012(2003).
DR   EMBL; AB085613; BAC66370.1; -.
DR   GO; GO:0008168; F:methyltransferase activity; IEA.
DR   GO; GO:0008174; F:mRNA methyltransferase activity; IEA.
DR   GO; GO:0003724; F:RNA helicase activity; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006396; P:RNA processing; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   InterPro; IPR000606; Viral_helicase1.
DR   InterPro; IPR002588; V_methyltransf.
DR   Pfam; PF01443; Viral_helicase1; 1.
DR   Pfam; PF01660; Vmethyltransf; 1.
DR   PROSITE; PS00626; RCC1_2; 1.
KW   Helicase; Methyltransferase; Polyprotein; Transferase.
SQ   SEQUENCE   1887 AA;  214910 MW;  94664A75D2691043 CRC64;

  Query Match             0.7%;  Score 9;  DB 12;  Length 1887;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        547 YYYLESQKL 555
              |||||||||
Db       1453 YYYLESQKL 1461


Search completed: September 10, 2004, 05:12:55
Job time : 56.5064 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 85.1587 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 6759
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6736	99.7	1291	2	AAY05814	Aay05814 Non-toxic
2	4201.5	62.2	811	5	ABG69079	Abg69079 Botulinum
3	2400.5	35.5	824	5	ABG69081	Abg69081 Botulinum
4	2386	35.3	450	3	AAY77136	Aay77136 Synthetic
5	2386	35.3	450	4	AAB04092	Aab04092 Botulism
6	2330	34.5	444	5	ABG69069	Abg69069 Botulinum
7	2317	34.3	462	2	AAW68397	Aaw68397 Clostridi
8	2287	33.8	436	5	ABG69078	Abg69078 Botulinum
9	2043	30.2	399	4	AAB04098	Aab04098 Botulism
10	1894	28.0	1291	2	AAW68392	Aaw68392 Clostridi
11	1853.5	27.4	1169	2	AAW56017	Aaw56017 Recombina
12	1824.5	27.0	1315	4	AAB61169	Aab61169 Clostridi
13	1769	26.2	1296	2	AAR95010	Aar95010 C. botuli
14	1763	26.1	1295	5	AAU99339	Aau99339 Clostridi
15	1589	23.5	399	3	AAY78982	Aay78982 C. botuli
16	1376.5	20.4	386	4	AAB04099	Aab04099 Botulism
17	1367	20.2	858	2	AAW56018	Aaw56018 Recombina
18	1348	19.9	852	5	ABG69077	Abg69077 Botulinum
19	1322	19.6	848	5	ABG69087	Abg69087 Botulinum
20	1255.5	18.6	1070	3	AAY93308	Aay93308 A mangane
21	1255.5	18.6	1095	3	AAY93311	Aay93311 A mangane
22	1247.5	18.5	848	4	AAB04082	Aab04082 Botulism
23	1245.5	18.4	858	5	ABG69085	Abg69085 Botulinum
24	1232.5	18.2	871	2	AAW56008	Aaw56008 Botulinum
25	1224.5	18.1	871	2	AAW56019	Aaw56019 Recombina
26	1224.5	18.1	871	2	AAW56007	Aaw56007 Recombina
27	1224	18.1	907	2	AAW56012	Aaw56012 Recombina
28	1220.5	18.1	875	2	AAW56009	Aaw56009 Recombina
29	1219.5	18.0	837	3	AAY77140	Aay77140 Native bo
30	1219.5	18.0	1067	3	AAY93307	Aay93307 A mangane
31	1219.5	18.0	1092	3	AAY93310	Aay93310 A mangane
32	1217.5	18.0	804	5	ABG69083	Abg69083 Botulinum
33	1216.5	18.0	873	2	AAW56016	Aaw56016 Recombina
34	1216.5	18.0	894	2	AAW56015	Aaw56015 Recombina
35	1212	17.9	953	2	AAW56011	Aaw56011 Recombina
36	1212	17.9	1013	2	AAW56013	Aaw56013 Recombina
37	1211	17.9	878	2	AAW56010	Aaw56010 Recombina
38	1209.5	17.9	861	5	ABG69075	Abg69075 Botulinum
39	1208	17.9	1059	3	AAY93309	Aay93309 A mangane
40	1208	17.9	1084	3	AAY93312	Aay93312 A mangane
41	1184	17.5	847	4	AAB04081	Aab04081 Botulism
42	1039.5	15.4	440	5	ABG69080	Abg69080 Botulinum
43	1039.5	15.4	441	5	ABG69070	Abg69070 Botulinum
44	915.5	13.5	451	4	AAB04093	Aab04093 Botulism
45	801.5	11.9	451	2	AAW68398	Aaw68398 Clostridi

ALIGNMENTS

RESULT 1
AAY05814
ID AAY05814 standard; protein; 1291 AA.
XX
AC AAY05814;
XX
DT 02-AUG-1999 (first entry)
XX
DE Non-toxic modified botulinum toxin rBoNT/C.
XX
KW Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery; mutant.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 229
FT /note= "His in native toxin"
FT Misc-difference 230
FT /note= "Glu in native toxin"
FT Misc-difference 233
FT /note= "His in native toxin"
XX
PN WO9920306-A1.
XX
PD 29-APR-1999.
XX
PF 16-OCT-1998; 98WO-US021897.
XX
PR 20-OCT-1997; 97US-00954302.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Simpson L, Kiyatkin N, Maksymowych A;
XX
DR WPI; 1999-302646/25.
DR N-PSDB; AAX25521.
XX
PT Modified toxin useful for systemic delivery of oral vaccines and
PT therapeutic agents.
XX
PS Example 1; Page 31-33; 37pp; English.
XX
CC The present sequence represents a modified serotype C botulinum toxin,
CC termed rBoNT/C, in which amino acids His-229, Glu-230 and His-233 of the
CC native sequence are substituted by Gly, Thr and Asn, respectively, i.e.
CC the zinc binding motif (see AAY05817) of the light chain holotoxin is
CC modified, resulting in loss of endoprotease activity. DNA coding for the

CC modified botulinum toxin (see AAX25521) was assembled from 3 separate
CC toxin fragments using PCR and site-directed mutagenesis. The modified
CC recombinant botulinum toxin maintains its ability to translocate from the
CC gut into the general circulation but is non-toxic. It can be used as an
CC oral vaccine for antigenic peptides including botulinum toxin (i.e. an
CC oral vaccine for botulism) or for the oral delivery of other therapeutic
CC agents to the general circulation
XX
SQ Sequence 1291 AA;

Query Match 99.7%; Score 6736; DB 2; Length 1291;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1288; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              |||||||||||||||||||||||||||||||||||||||||||||||  ||:|||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMGTLNNAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 2
ABG69079
ID ABG69079 standard; protein; 811 AA.
XX
AC ABG69079;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #13.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98549.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 142-144; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum

```
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 811 AA;

  Query Match          62.2%;  Score 4201.5;  DB 5;  Length 811;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-225;
  Matches  811;  Conservative    0;  Mismatches    0;  Indels    7;  Gaps    1;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 60

Qy         62 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 120

Qy        122 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 180

Qy        182 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 240

Qy        242 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 300

Qy        302 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 360

Qy        362 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 420

Qy        422 RKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKD 481
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        421 RKVNPENMLYLFTKFC-------SLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKD 473

Qy        482 INEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNV 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 INEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNV 533

Qy        542 DYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMW 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 DYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMW 593

Qy        602 ANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLE 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        594 ANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLE 653

Qy        662 AFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFN 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        654 AFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFN 713

Qy        722 NISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINK 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        714 NISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINK 773

Qy        782 FIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLID 819
              |||||||||||||||||||||||||||||||||||||
Db        774 FIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLID 811


RESULT 3
ABG69081
ID   ABG69081 standard; protein; 824 AA.
XX
AC   ABG69081;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #15.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98551.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 52; Page 147-149; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 824 AA;

  Query Match          35.5%;  Score 2400.5;  DB 5;  Length 824;
  Best Local Similarity   57.6%;  Pred. No. 2.8e-125;
  Matches  476;  Conservative  115;  Mismatches  229;  Indels    7;  Gaps    3;

Qy          5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
              : :|||||:: :||||    | |    | ||| || | ||||:||| ::||:|:||||
Db          4 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNTWVIPERFSSDTNPSLSKPPRP 63

Qy         65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124
              ||    ||||:|||||  |||||| |||||||| |:||::||  |    || |:::||
Db         64 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 123
```

```
Qy        125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184
Db        124 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 183

Qy        185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244
Db        184 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 243

Qy        245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304
Db        244 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEII-QIERSQLREKALGHYKDIAKRL 302

Qy        305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364
Db        303 NNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSEVVYS 362

Qy        365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424
Db        363 SQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPALQKL 422

Qy        425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484
Db        423 SSESVVDLFTKVCL-----RLTKNSRDDSTCIKVKNNRLPYVADKDSISQEIFENKIITD 477

Qy        485 ETEVIYYPDNVSVDQVILSKNTSEHGQL-DLLYPSIDSESEILPGENQVFYDNRTQNVDY 543
Db        478 ETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFYDDITKYVDY 537

Qy        544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603
Db        538 LNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 597

Qy        604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663
Db        598 EVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFATAGVAFLLEGF 657

Qy        664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723
Db        658 PEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHI 717

Qy        724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783
Db        718 NYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 777

Qy        784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVD 830
Db        778 RECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVD 824
```

RESULT 4
AAY77136
ID AAY77136 standard; protein; 450 AA.
XX
AC AAY77136;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype C; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87214.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 24; Page 41-42; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 450 AA;

Query Match 35.3%; Score 2386; DB 3; Length 450;
Best Local Similarity 100.0%; Pred. No. 8.1e-125;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        843 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 902
Db          2 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 61

Qy        903 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962
Db         62 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

Qy        963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 1022
Db        122 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 181

Qy       1023 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
Db        182 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241

Qy       1083 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 1142
Db        242 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 301

Qy       1143 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202
Db        302 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361

Qy       1203 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1262
Db        362 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 421

Qy       1263 YLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

```
Db        422 YLVPTVKQGNYASLLESTSTHWGFVPVSE 450


RESULT 5
AAB04092
ID   AAB04092 standard; protein; 450 AA.
XX
AC   AAB04092;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype C).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;  99US-0133865P.
PR   12-MAY-1999;  99US-0133866P.
PR   12-MAY-1999;  99US-0133867P.
PR   12-MAY-1999;  99US-0133868P.
PR   12-MAY-1999;  99US-0133869P.
PR   12-MAY-1999;  99US-0133873P.
PR   29-JUL-1999;  99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54486.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 5b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 450 AA;

  Query Match             35.3%;  Score 2386;  DB 4;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-125;
  Matches  449;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        843 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 902
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 TIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLN 61

Qy        903 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 PIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

Qy        963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 1022
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKI 181

Qy       1023 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 YINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241

Qy       1083 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 1142
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 KDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDF 301

Qy       1143 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 NEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361

Qy       1203 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1262
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 421

Qy       1263 YLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db        422 YLVPTVKQGNYASLLESTSTHWGFVPVSE 450


RESULT 6
ABG69069
ID   ABG69069 standard; protein; 444 AA.
XX
AC   ABG69069;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #3.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98539.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 121-122; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
```

```
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 444 AA;

  Query Match             34.5%;  Score 2330;  DB 5;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 1e-121;
  Matches  444;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGR 444
              ||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGR 444


RESULT 7
AAW68397
ID   AAW68397 standard; protein; 462 AA.
XX
AC   AAW68397;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type C1 toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW   botulism; BotC.
XX
OS   Clostridium botulinum; serotype C1 strain Stockholm.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30588.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 45; Page 339-341; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC   DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC   express BotC soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             34.3%;  Score 2317;  DB 2;  Length 462;
  Best Local Similarity   99.8%;  Pred. No. 5.9e-121;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85

Qy        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145

Qy        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205

Qy       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265

Qy       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325

Qy       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385

Qy       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445

Qy       1275 SLLESTSTHWGFVPVSE 1291
              |||||||||||||||||
Db        446 SLLESTSTHWGFVPVSE 462


RESULT 8
```

```
ABG69078
ID   ABG69078 standard; protein; 436 AA.
XX
AC   ABG69078;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #12.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98548.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 140-141; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 436 AA;

  Query Match            33.8%;  Score 2287;  DB 5;  Length 436;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-119;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 60

Qy         62 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 120

Qy        122 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 180

Qy        182 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 240

Qy        242 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 300

Qy        302 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 360

Qy        362 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 420

Qy        422 RKVNPENMLYLFTKFC 437
              ||||||||||||||||
Db        421 RKVNPENMLYLFTKFC 436


RESULT 9
AAB04098
ID   AAB04098 standard; protein; 399 AA.
XX
AC   AAB04098;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain N-terminal sequence (serotype C).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54494.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 6; Fig 13b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
```

CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 399 AA;

Query Match 30.2%; Score 2043; DB 4; Length 399;
Best Local Similarity 100.0%; Pred. No. 8.8e-106;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        445 SLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSK 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSK 61

Qy        505 NTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTF 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 NTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTF 121

Qy        565 TRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 181

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 241

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 301

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 361

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQN 842
              |||||||||||||||||||||||||||||||||||||
Db        362 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQN 399
```

RESULT 10
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30579.

XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 291-296; 428pp; English.
XX
CC This is the amino acid sequence of the type B toxin of Clostridium
CC botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC B toxin has been expressed as a histidine-tagged protein in Escherichia
CC coli. The invention relates to C. botulinum recombinant toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of the recombinant proteins are E. coli, insect cells and
CC yeast cells. The recombinant proteins are used as immunogens for the
CC production of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1291 AA;

Query Match 28.0%; Score 1894; DB 2; Length 1291;
Best Local Similarity 33.9%; Pred. No. 8.2e-97;
Matches 462; Conservative 248; Mismatches 504; Indels 150; Gaps 39;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||:|||||||:||:||  ||: ::           |||:||  ||:||:|::    | : |
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
               |          ||         ||||:||:|:  |: ||: :|||| || |: :||:|:  :
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
              ||: |:   |:  |: ::   :|:          :   |  ::|| ||   : : ||
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
                : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
              | :| ||||       :||::         |: |    ::  |:| |||    :| |
Db        233 IHVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIITPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
              |   :| |  :| |  ||| :     |  |  |  || |   ||:|| :| |: :::
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
              | :||  |   ||| | |: | :: |  | |:   ||   |:||: :| |: ||||
Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
              ::   : ||| : |  |  ::: |::     : | |::     :       | :  | | ||
Db        404 DMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES 523
              || | :    |:  : |   |:  |  ::  ::::||  :|    :::|  || ::||
Db        455 FFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELIL--DTDLISKIEL--PSENTES 510

Qy        524 ----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
                       |  |  ::| | |         : || ||    :: | : | | ::||
Db        511 LTDFNVDVPVYEKQPAIKKIFTDENT-------IFQYLYSQTFLLDIRDISLTSSFDDAL 563

Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
              | |||::|       ||||   |: |||  |   :| ||      : :|:|||:|:| |
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
              :|||| |||: |   :|||  || : | :||||  ||  || :|||:: | :  :|:|||
Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIK 680
```

```
Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
Db        681 TIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
Db        741 YSEKEKSNINIDFNDINSKLNEGINQAIHNINNFINGCSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
Db        801 TLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
Db        861 LNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA--NSKIRVTQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
Db        917 IIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLI 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
Db        977 DINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKD 1100
Db       1036 ANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKD 1086

Qy       1101 YWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN-------DFNEGYKII 1149
Db       1087 FWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFI 1146

Qy       1150 IKRIRGNT---NDTRVRGGDILYFD-MTINNK----AYNLFMKNETMYADNHSTEDIYAI 1201
Db       1147 IRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE---------EKLFLA 1196

Qy       1202 GLREQTKDINDNIIFQIQPMNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRLGG---- 1256
Db       1197 PISDSDEFYN---TIQIKEYHEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYESGIVFE 1251

Qy       1257 ---------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db       1252 EYKDYFCISKWY-----LKEVKRKPYNLKL---GCNWQFIPKDE 1287
```

```
RESULT 11
AAW56017
ID   AAW56017 standard; protein; 1169 AA.
XX
AC   AAW56017;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH728/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW   detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB002273.
XX
PR   23-AUG-1996;   96GB-00017671.
PR   13-DEC-1996;   96GB-00025996.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Shone CC,  Quinn CP,  Foster KA;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26289.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT   immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS   Example 2; Page 91-94; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from the
CC   present invention. The present invention describes recombinant neurotoxin
CC   proteins which comprise a first and second domain, where the first domain
CC   is adapted to cleave one or more vesicle or plasma-membrane associated
CC   proteins essential to exocytosis, and where the second domain is adapted:
CC   (a) to translocate the protein into a cell; (b) to increase the
CC   solubility of the protein compared to the solubility of the first domain
CC   on its own, or (c) both to translocate the protein into a cell and to
CC   increase the solubility of the protein compared to the solubility of the
CC   first domain on its own, the protein being free of clostridial neurotoxin
CC   (CN) and free of CN precursor that can be converted into toxin by
CC   proteolytic action. The recombinant proteins can be used as therapeutic
CC   agents for targeting cells expressing a relevant substrate. The products
CC   can also be used as immunogens and as non-toxic standards for the
CC   assessment and development of in vitro assays for the detection of
CC   functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC   environmental samples
XX
SQ   Sequence 1169 AA;

  Query Match             27.4%;  Score 1853.5;  DB 2;  Length 1169;
  Best Local Similarity   35.2%;  Pred. No. 1.3e-94;
  Matches  431;  Conservative  230;  Mismatches  453;  Indels  109;  Gaps   30;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db        233 IHVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIITPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
Db        404 DMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES 523
Db        455 FFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELIL--DTDLISKIEL--PSENTES 510

Qy        524 ----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
Db        511 LTDFNVDVPVYEKQPAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDAL 563
```

```
Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIK 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
Db        681 TIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
Db        741 YSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
Db        801 TLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
Db        861 LNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA--NSKIRVTQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
Db        917 IIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLI 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
Db        977 DINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKD 1100
Db       1036 ANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKD 1086

Qy       1101 YWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN-------DFNEGYKII 1149
Db       1087 FWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFI 1146

Qy       1150 IKRIRGNT---NDTRVRGGDILY 1169
Db       1147 IRR-KSNSQSINDDIVRKEDYIY 1168
```

RESULT 12
AAB61169
ID AAB61169 standard; protein; 1315 AA.
XX
AC AAB61169;
XX
DT 02-APR-2001 (first entry)
XX
DE Clostridium tetani TeNT.
XX
KW Clostridium tetani; TeNT; tetanus toxin; antibacterial; vaccine;
KW TeNT fragment C; infection.
XX
OS Clostridium tetani.
XX
PN WO200100839-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-GB002428.
XX
PR 25-JUN-1999; 99GB-00014861.
XX
PA (UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI Fairweather NF, Sinha K;
XX
DR WPI; 2001-123014/13.
XX
PT New polypeptide, useful for treating infections of Clostridium tetani,
PT comprises tetanus toxin fragment with a mutation in a loop region,.
XX
PS Disclosure; Page 39; 43pp; English.
XX
CC The present sequence is given in a specification relating to a novel
CC polypeptide comprising tetanus toxin (TeNT) fragment C or its immunogenic
CC fragment, containing a mutation in a loop region. The mutation results in
CC a reduction in the binding of TeNT fragment C or its immunogenic fragment
CC to gangliosides and primary motoneurones, and/or a reduction in the
CC ability of TeNT fragment C or its immunogenic fragment to undergo
CC retrograde transport. The polypeptide is useful for treating, preventing
CC and reducing the susceptibility to Clostridium tetani infection in a
CC human or animal, and also for producing antibodies which recognise groups
CC within TeNT polypeptides. Antibody produced against the polypeptide is
CC also useful for treating Clostridium tetani infection
XX
SQ Sequence 1315 AA;

Query Match 27.0%; Score 1824.5; DB 4; Length 1315;
Best Local Similarity 32.6%; Pred. No. 6.3e-93;
Matches 448; Conservative 252; Mismatches 527; Indels 147; Gaps 33;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        120 NNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKL---T 176
Db        121 SYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDN 180

Qy        177 NNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMH 236
Db        181 KNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLH 240

Qy        237 NLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDY 296
Db        241 GLYGMQVSSHEIIPS--KQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLND 298

Qy        297 YRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQ 356
Db        299 YKAIANKLSQVTSCNDPNID--IDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMY 356

Qy        357 IFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLFMGQNL 415
Db        357 GFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNM 416

Qy        416 SRNP-ALRKVNPENMLYLFTKFCHKAID----GRSLYNKTLDCREL------LVKNTDLP 464
Db        417 RVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLT 476

Qy        465 FIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESE 524
Db        477 FIAEKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNL----QSKITLPN-DRTTP 531

Qy        525 ILPG-------ENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAK 577
Db        532 VTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLGRITMTNSVDDALINSTK 591

Qy        578 VYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637
Db        592 IYSYFPSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALN 651

Qy        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQR 697
Db        652 IVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKR 711
```

```
Qy        698 IKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENI 757
Db        712 YEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQI 771

Qy        758 KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINL 817
Db        772 ADEINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQY 831

Qy        818 IDSHN-IILVGEVDKLKAKVNNSFQNTIPF----NIFSYTNNSLLKDIINEYFNNINDSK 872
Db        832 IKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVI------LKKST 885

Qy        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGKVIVTQ 925
Db        886 ILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSEVIVHK 937

Qy        926 NENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWSIGIIS 974
Db        938 AMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKG 997

Qy        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032
Db        998 NNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSA 1056

Qy       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092
Db       1057 EITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEKLYTSY 1108

Qy       1093 QYTNVVKDYWGNDLRYNKEYYMVNI-------------DYL---NRYMYANSRQIVFNTR 1136
Db       1109 LSITFLRDFWGNPLRYDTEYYLIPVASSSKDVGLKNITDYMYLTNAPSYTNGKLNIYYRR 1168

Qy       1137 RNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLFMKNET 1187
Db       1169 LYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDR 1223

Qy       1188 MYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI 1247
Db       1224 ILRVGYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL--KLYDDKNAS-LGLV 1268

Qy       1248 GTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db       1269 GTHNGQIGNDPNRDILIASNWY-FNHLKDKI----------LGCDWYFVPTDE 1310
```

RESULT 13
AAR95010
ID AAR95010 standard; protein; 1296 AA.
XX
AC AAR95010;
XX
DT 09-JUL-1996 (first entry)
XX
DE C. botulinum type A neurotoxin.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.
XX
OS Clostridium botulinum.
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29244.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 4; Page 344-350; 434pp; English.
XX
CC Clostridium botulinum type A neurotoxin (AAR95010) is processed to form a
CC dimer composed of a light and a heavy chain. It is the product of the
CC type A neurotoxin gene (AAT29244). The 50 kDa C-terminal portion of the
CC heavy chain, or C fragment (see also AAR95008), was produced using a
CC synthetic gene (AAT29245) having codon usage altered to improve
CC expression in Escherichia coli. Fusion proteins of the type A toxin or C
CC fragment, with e.g. maltose binding protein or polyhistidine affinity tag
CC (see also AAR95008), are used to generate neutralising antitoxins and in
CC vaccine compsns
XX
SQ Sequence 1296 AA;

Query Match 26.2%; Score 1769; DB 2; Length 1296;
Best Local Similarity 32.8%; Pred. No. 7.6e-90;
Matches 443; Conservative 243; Mismatches 543; Indels 122; Gaps 34;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505
Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
```

```
Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744

Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804

Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864

Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869

Qy        865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKV 921

Db        870 IKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KI 924

Qy        922 IVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLV 978

Db        925 EVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEII 984

Qy        979 FTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKEL 1037

Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1038 TGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNV 1097

Db       1045 GNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR-RNNNDFNE 1144

Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1145 GYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLR 1204

Db       1157 GTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSA 1208

Qy       1205 EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYL 1264

Db       1209 LEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNI 1258

Qy       1265 VPTVKQGNYASLLESTS----THWGFVPVSE 1291

Db       1259 AKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289
```

RESULT 14
AAU99339
ID AAU99339 standard; protein; 1295 AA.
XX
AC AAU99339;
XX
DT 07-OCT-2002 (first entry)
XX
DE Clostridium botulinum botulism A toxin protein.
XX
KW Botulism A toxin; neuroprotective; muscular; BoTOX; tetanus toxin; TeTOX;
KW protease cleavage; toxicity; neurotoxin; poison; neural cell; toxin;
KW post-translational modification; light chain; heavy chain;
KW spastic condition; stabismus; bephlarospasm; hemifacial spasm;
KW brain injury; spinal cord injury; stroke; multiple sclerosis;
KW cerebral palsy; BoNT/A.
XX
OS Clostridium botulinum.
XX

FH	Key	Location/Qualifiers
FT	Region	1. .437
FT		/note= "Light chain"
FT	Active-site	222
FT		/note= "Forms an active site with Glu223, His226, Glu261,
FT		Tyr365"
FT	Active-site	223
FT		/note= "Forms an active site with His222, His226, Glu261,
FT		Tyr365"
FT	Active-site	226
FT		/note= "Forms an active site with His222, Glu223, Glu261,
FT		Tyr365"
FT	Active-site	261
FT		/note= "Forms an active site with His222, Glu223, His226,
FT		Tyr365"
FT	Active-site	365
FT		/note= "Forms an active site with His222, Glu223, His226,
FT		Glu261"
FT	Region	437. .448
FT		/note= "Linker region"
FT	Region	449. .1295
FT		/note= "Heavy chain"
FT	Misc-difference	1161
FT		/label= OTHER
FT		/note= "OTHER = O, which is not defined in the
FT		specification"

XX
PN WO200244199-A2.
XX
PD 06-JUN-2002.
XX
PF 27-NOV-2001; 2001WO-US045059.
XX
PR 29-NOV-2000; 2000US-00726949.
XX
PA (ALLR) ALLERGAN SALES INC.
XX
PI Lin W, Aoki KR, Steward LE;
XX
DR WPI; 2002-557531/59.
XX
PT Novel modified botulism toxin or tetanus toxin comprising a protease
PT cleavage site, is useful for treating conditions benefited by neurotoxin
PT activity.
XX
PS Disclosure; Fig 1; 19pp; English.
XX
CC The invention discloses modified botulism toxin (BoTOX) or tetanus toxin
CC (TeTOX) which contain new protease cleavage sites. These sites are
CC created in regions where the cleavage site is specifically susceptible to
CC cleavage by a protease that is present in effective levels only in a
CC tissue where toxicity is undesirable and where the cleavage site itself
CC does not inactivate the toxin. BoTOX and TeTOX are neurotoxins which are
CC highly potent and specific poisons of neural cells. The toxins are
CC synthesised in vivo as single chains, which are not toxic, but become
CC active when nicked, in a post-translational modification, to form
CC separate light and heavy chains which are linked through a disulphide
CC bond. The modified toxins are useful to treat conditions benefited by
CC neurotoxin activity (e.g. spastic conditions, including stabismus,
CC bephlarospasm and hemifacial spasm, brain injury, spinal cord injury,
CC stroke, multiple sclerosis and cerebral palsy) by administering the
CC toxins for their localised production. The advantage of the toxins are
CC that they are deactivated in tissues where toxic activity is undesirable
CC and activated at desired targets. The sequence presented is the
CC Clostridium botulinum botulism A toxin (BoNT/A) protein
XX
SQ Sequence 1295 AA;

Query Match 26.1%; Score 1763; DB 5; Length 1295;
Best Local Similarity 32.7%; Pred. No. 1.6e-89;
Matches 441; Conservative 244; Mismatches 543; Indels 122; Gaps 34;

```
Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNLN 59

Db          1 PFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN 59

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119

Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119
```

```
Qy       120 NNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF--- 173
Db       120 -----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGHE 170

Qy       174 --KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPILI 226
Db       171 VLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAVT 219

Qy       227 LMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSAR 286
Db       220 LAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQE 278

Qy       287 KYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGEV 340
Db       279 NEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGKF 330

Qy       341 TVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFN 400
Db       331 SVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFN 390

Qy       401 IPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTLD 452
Db       391 LRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALN 450

Qy       453 CRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KNT 506
Db       451 DLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDNE 510

Qy       507 SEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTR 566
Db       511 PENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTN 569

Qy       567 SIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625
Db       570 SVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADI 629

Qy       626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685
Db       630 TIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVL 689

Qy       686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745
Db       690 TVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ 749

Qy       746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805
Db       750 YNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLED 809

Qy       806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865
Db       810 FDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYI 869

Qy       866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVI 922
Db       870 KNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KIE 924

Qy       923 VTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVF 979
Db       925 VILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIW 984

Qy       980 TLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELT 1038
Db       985 DLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG 1044

Qy      1039 GINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVV 1098
Db      1045 NIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1096

Qy      1099 KDYWGNDLRYNKEYYMVNIDYLNRYMYANS-----------RQIVFNTR-RNNNDFNEG 1145
Db      1097 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy      1146 YKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLRE 1205
Db      1157 TKFIKKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQVVVFKILSAL 1208

Qy      1206 QTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLV 1265
Db      1209 EIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNIA 1258

Qy      1266 PTVKQGNYASLLESTS----THWGFVPVSE 1291
Db      1259 KLVASNWYNRQIERSSRTLGCSWEFIPVDD 1288
```

RESULT 15
AAY78982
ID AAY78982 standard; protein; 399 AA.
XX
AC AAY78982;
XX
DT 20-JUN-2000 (first entry)
XX
DE C. botulinum type D toxin amino acid sequence.
XX
KW Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
KW protect.
XX
OS Clostridium botulinum.
XX
PN WO200005252-A1.
XX
PD 03-FEB-2000.
XX
PF 20-JUL-1999; 99WO-IB001301.
XX
PR 22-JUL-1998; 98ZA-00006538.
XX
PA (AGRI-) AGRIC RES COUNCIL.
XX
PI De Bruyn EE, Botha AD;
XX
DR WPI; 2000-205375/18.
DR N-PSDB; AAZ98630, AAZ98631.
XX
PT Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin type
PT D, useful in vaccines for protection against botulism, comprises at least
PT one amino acid mutation not present in the wild type D neurotoxins.
XX
PS Claim 3, 4; Page 54-57; 66pp; English.
XX
CC This sequence represents the amino acid sequence of a synthetic non-toxic
CC immunogenic derivative of Clostridium botulinum type D toxin (BoNT).
CC Botulinum neurotoxin causes botulism poisoning in cattle and sheep, and
CC usually results in the death of the affected or poisoned animal. The non-
CC toxic immunogenic fragments of the C. botulinum neurotoxin are useful in
CC vaccines to protect animals (e.g. humans, cattle, sheep, pigs) against
CC BoNT type D poisoning. The non-toxic fragments can be produced relatively
CC simply and inexpensively (specifically by fermentation techniques). As
CC the fragments are not toxic the risk to production staff is reduced
XX
SQ Sequence 399 AA;

Query Match 23.5%; Score 1589; DB 3; Length 399;
Best Local Similarity 76.0%; Pred. No. 1.7e-80;
Matches 307; Conservative 33; Mismatches 56; Indels 8; Gaps 4;

```
Qy       891 AEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWV 950
Db         1 AEVRVEGNVQLNPIFPFDFKLGSSGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWV 60

Qy       951 SNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFV 1010
Db        61 SNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFV 120
```

```
Qy       1011 TVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWI 1070
              |:| |||||| ||||||||||||||||||||||||||::||||:|||||||||||||||
Db        121 TITTNMMGNMMIYINGKLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWI 180

Qy       1071 RDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQ 1130
              |||||||||| ||||||||||||||||||||||||:|||||:|::|:||||
Db        181 RDFYIFAKELDDKDINILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNG 240

Qy       1131 IVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA 1190
              ||||||:|||||||||||||||||||||||| ::|||: ||:|| |:|       ||
Db        241 IVFNTRKNNNDFNEGYKIIIKRIRGNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYK 295

Qy       1191 DNHST-EDIYAIGLREQTKD-INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIG 1248
                : :   |:  :|  :|  | |     | ||| |   |||||:| |:    |  || |||
Db        296 PSRNLGTDLVPLGALDQPMDEIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIG 355

Qy       1249 TYRFRLGGD-WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              :| |:|| | |: | ||:| :|  :|||||||||||| ||| ||
Db        356 SYSFKLGDDYWFNHEYLIPVIKIEHYASLLESTSTHWVFVPASE 399
```

Search completed: September 10, 2004, 03:33:38
Job time : 89.1587 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 27.4018 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 6759
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

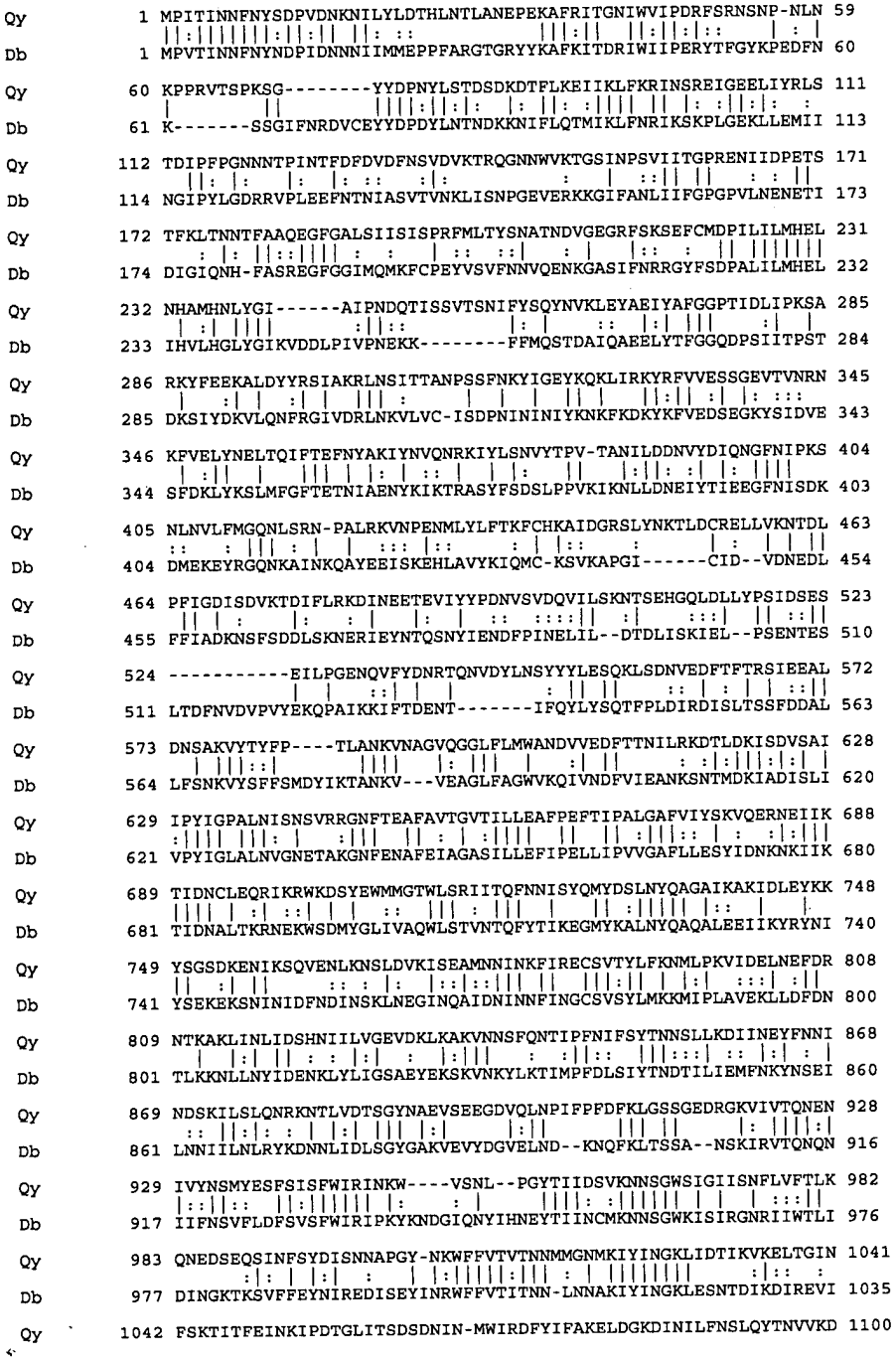
SUMMARIES

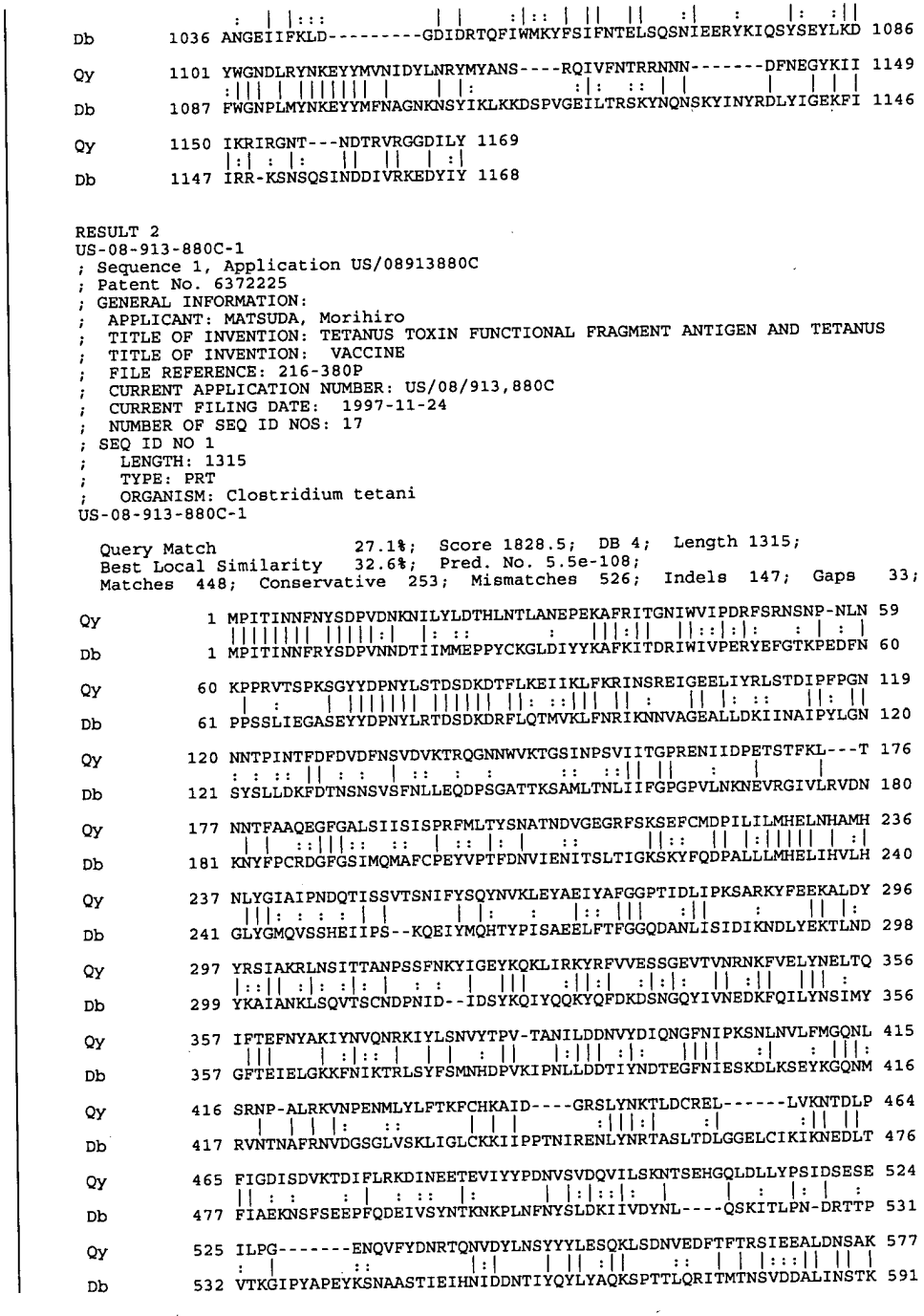
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1853.5	27.4	1169	4	US-09-255-829-20	Sequence 20, Appl
2	1828.5	27.1	1315	4	US-08-913-880C-1	Sequence 1, Appli
3	1769	26.2	1296	1	US-08-480-604A-28	Sequence 28, Appl
4	1769	26.2	1296	2	US-08-405-496A-28	Sequence 28, Appl
5	1769	26.2	1296	3	US-08-915-136-28	Sequence 28, Appl
6	1769	26.2	1296	4	US-09-084-517-28	Sequence 28, Appl
7	1367	20.2	858	4	US-09-255-829-22	Sequence 22, Appl
8	1367	20.2	858	4	US-09-255-829-29	Sequence 29, Appl
9	1232.5	18.2	871	4	US-09-255-829-8	Sequence 8, Appli
10	1224.5	18.1	871	4	US-09-255-829-2	Sequence 2, Appli
11	1224.5	18.1	871	4	US-09-255-829-26	Sequence 26, Appl
12	1224	18.1	907	4	US-09-255-829-16	Sequence 16, Appl
13	1220.5	18.1	875	4	US-09-255-829-10	Sequence 10, Appl
14	1216.5	18.0	873	4	US-09-255-829-6	Sequence 6, Appli
15	1216.5	18.0	894	4	US-09-255-829-4	Sequence 4, Appli
16	1212	17.9	953	4	US-09-255-829-14	Sequence 14, Appl
17	1212	17.9	1013	4	US-09-255-829-18	Sequence 18, Appl
18	1211	17.9	878	4	US-09-255-829-12	Sequence 12, Appl
19	1158.5	17.1	865	4	US-08-913-880C-13	Sequence 13, Appl
20	1158.5	17.1	866	4	US-08-913-880C-12	Sequence 12, Appl
21	1158.5	17.1	874	4	US-08-913-880C-11	Sequence 11, Appl
22	1158.5	17.1	875	4	US-08-913-880C-10	Sequence 10, Appl
23	1146.5	17.0	862	4	US-08-913-880C-14	Sequence 14, Appl
24	1145.5	16.9	853	4	US-08-913-880C-17	Sequence 17, Appl
25	1145.5	16.9	858	4	US-08-913-880C-16	Sequence 16, Appl
26	1145.5	16.9	860	4	US-08-913-880C-15	Sequence 15, Appl
27	710.5	10.5	548	4	US-09-255-829-24	Sequence 24, Appl
28	557.5	8.2	438	1	US-08-480-604A-23	Sequence 23, Appl
29	557.5	8.2	438	2	US-08-405-496A-23	Sequence 23, Appl
30	557.5	8.2	438	3	US-08-915-136-23	Sequence 23, Appl
31	557.5	8.2	438	4	US-09-084-517-23	Sequence 23, Appl
32	557.5	8.2	462	1	US-08-480-604A-26	Sequence 26, Appl
33	557.5	8.2	462	2	US-08-405-496A-26	Sequence 26, Appl
34	557.5	8.2	462	3	US-08-915-136-26	Sequence 26, Appl
35	557.5	8.2	462	4	US-09-084-517-26	Sequence 26, Appl
36	424	6.3	618	1	US-08-668-381A-5	Sequence 5, Appli
37	420	6.2	452	1	US-07-618-312A-2	Sequence 2, Appli
38	420	6.2	452	1	US-07-618-312A-4	Sequence 4, Appli
39	420	6.2	452	1	US-08-280-228-2	Sequence 2, Appli
40	420	6.2	452	1	US-08-280-228-4	Sequence 4, Appli
41	397	5.9	452	1	US-08-110-786A-8	Sequence 8, Appli
42	280	4.1	2184	4	US-09-417-485D-6	Sequence 6, Appli
43	263	3.9	207	3	US-08-446-114A-21	Sequence 21, Appl
44	263	3.9	2710	1	US-08-480-604A-6	Sequence 6, Appli
45	263	3.9	2710	2	US-08-405-496A-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-20

  Query Match            27.4%;  Score 1853.5;  DB 4;  Length 1169;
  Best Local Similarity  35.2%;  Pred. No. 1.2e-109;
  Matches  431;  Conservative  230;  Mismatches  453;  Indels  109;  Gaps  30;
```





RESULT 2
US-08-913-880C-1
; Sequence 1, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
; APPLICANT: MATSUDA, Morihiro
; TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
; TITLE OF INVENTION: VACCINE
; FILE REFERENCE: 216-380P
; CURRENT APPLICATION NUMBER: US/08/913,880C
; CURRENT FILING DATE: 1997-11-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-08-913-880C-1

Query Match 27.1%; Score 1828.5; DB 4; Length 1315;
Best Local Similarity 32.6%; Pred. No. 5.5e-108;
Matches 448; Conservative 253; Mismatches 526; Indels 147; Gaps 33;

```
Qy        578 VYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637
Db        592 IYSYFPSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALN 651

Qy        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQR 697
Db        652 IVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKR 711

Qy        698 IKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENI 757
Db        712 YEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQI 771

Qy        758 KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINL 817
Db        772 ADEINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQY 831

Qy        818 IDSHN-IILVGEVDKLKAKVNNSFQNTIPF----NIFSYTNNSLLKDIINEYFNNINDSK 872
Db        832 IKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVI------LKKST 885

Qy        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGKVIVTQ 925
Db        886 ILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSEVIVHK 937

Qy        926 NENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWSIGIIS 974
Db        938 AMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKG 997

Qy        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032
Db        998 NNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSA 1056

Qy       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092
Db       1057 EITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEKLYTSY 1108

Qy       1093 QYTNVVKDYWGNDLRYNKEYYMVNI-------------DYL---NRYMYANSRQIVFNTR 1136
Db       1109 LSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRR 1168

Qy       1137 RNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLFMKNET 1187
Db       1169 LYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDR 1223

Qy       1188 MYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI 1247
Db       1224 ILRVGYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL--KLYDDKNAS-LGLV 1268

Qy       1248 GTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db       1269 GTHNGQIGNDPNRDILIASNWY-FNHLKDKI----------LGCDWYFVPTDE 1310
```

```
RESULT 3
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match           26.2%;  Score 1769;  DB 1;  Length 1296;
  Best Local Similarity  32.8%;  Pred. No. 3.3e-104;
  Matches  443;  Conservative  243;  Mismatches  543;  Indels  122;  Gaps   34;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
```

```
Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451

Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505

Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565

Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624

Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684

Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744

Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNININKFIRECSVTYLFKNMLPKVIDELN 804

Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864

Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869

Qy        865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKV 921

Db        870 IKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KI 924

Qy        922 IVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLV 978

Db        925 EVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEII 984

Qy        979 FTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKEL 1037

Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1038 TGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNV 1097

Db       1045 GNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR-RNNNDFNE 1144

Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1145 GYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLR 1204

Db       1157 GTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSA 1208

Qy       1205 EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYL 1264

Db       1209 LEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNI 1258

Qy       1265 VPTVKQGNYASLLESTS----THWGFVPVSE 1291

Db       1259 AKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289


RESULT 4
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match           26.2%;  Score 1769;  DB 2;  Length 1296;
  Best Local Similarity  32.8%;  Pred. No. 3.3e-104;
  Matches  443;  Conservative  243;  Mismatches  543;  Indels  122;  Gaps   34;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58

Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118

Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173

Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225

Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285

Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339

Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330
```

```
Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505
Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864
Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869

Qy        865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKV 921
Db        870 IKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KI 924

Qy        922 IVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLV 978
Db        925 EVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEII 984

Qy        979 FTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKEL 1037
Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1038 TGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNV 1097
Db       1045 GNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR-RNNNDFNE 1144
Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1145 GYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLR 1204
Db       1157 GTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSA 1208

Qy       1205 EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYL 1264
Db       1209 LEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNI 1258

Qy       1265 VPTVKQGNYASLLESTS----THWGFVPVSE 1291
Db       1259 AKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289


RESULT 5
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match           26.2%;  Score 1769;  DB 3;  Length 1296;
  Best Local Similarity  32.8%;  Pred. No. 3.3e-104;
  Matches  443;  Conservative  243;  Mismatches  543;  Indels  122;  Gaps  34;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
```

```
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225

Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285

Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339

Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399

Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451

Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505

Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565

Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624

Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684

Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744

Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804

Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864

Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869

Qy        865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKV 921

Db        870 IKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KI 924

Qy        922 IVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLV 978

Db        925 EVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEII 984

Qy        979 FTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKEL 1037

Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1038 TGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNV 1097

Db       1045 GNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR-RNNNDFNE 1144

Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1145 GYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLR 1204

Db       1157 GTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSA 1208

Qy       1205 EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYL 1264

Db       1209 LEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNI 1258

Qy       1265 VPTVKQGNYASLLESTS----THWGFVPVSE 1291

Db       1259 AKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289
```

RESULT 6
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-28

Query Match 26.2%; Score 1769; DB 4; Length 1296;
Best Local Similarity 32.8%; Pred. No. 3.3e-104;
Matches 443; Conservative 243; Mismatches 543; Indels 122; Gaps 34;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58

Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59
```

```
Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505
Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864
Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869

Qy        865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKV 921
Db        870 IKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KI 924

Qy        922 IVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLV 978
Db        925 EVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEII 984

Qy        979 FTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKEL 1037
Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1038 TGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNV 1097
Db       1045 GNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR-RNNNDFNE 1144
Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1145 GYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLR 1204
Db       1157 GTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSA 1208

Qy       1205 EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYL 1264
Db       1209 LEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNI 1258

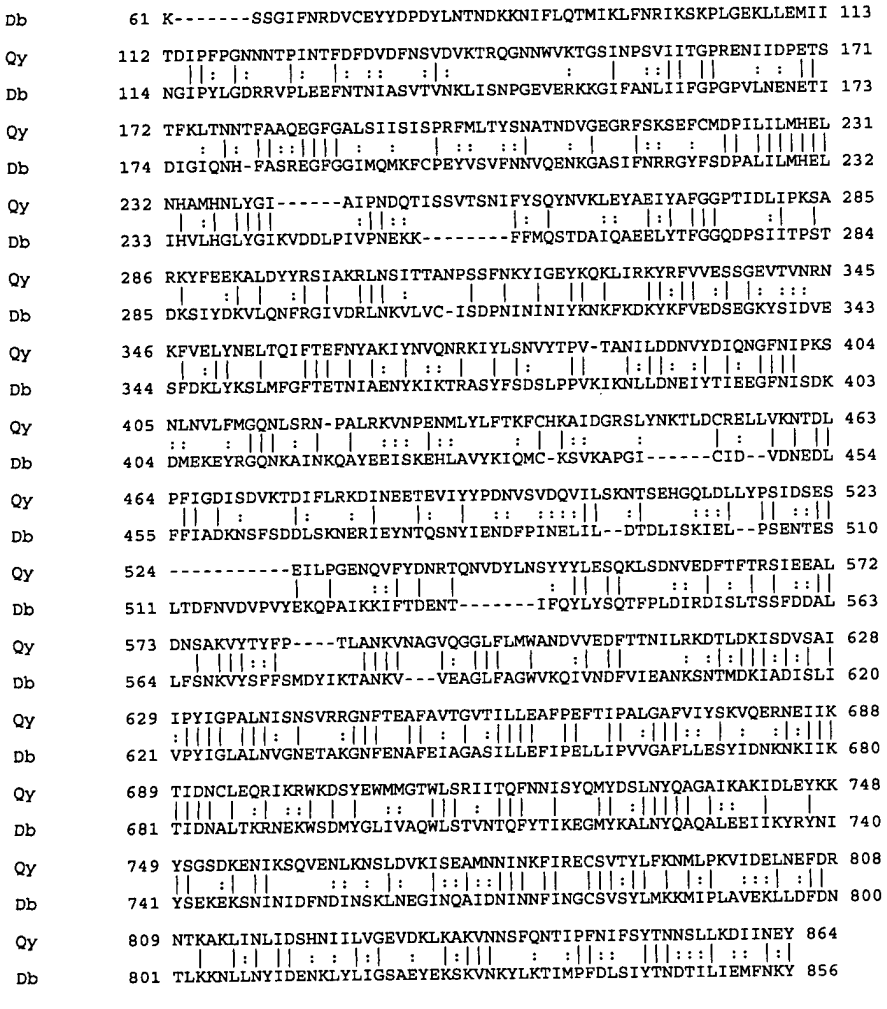
Qy       1265 VPTVKQGNYASLLESTS----THWGFVPVSE 1291
Db       1259 AKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289
```

RESULT 7
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-22

Query Match 20.2%; Score 1367; DB 4; Length 858;
Best Local Similarity 34.5%; Pred. No. 7.2e-79;
Matches 309; Conservative 175; Mismatches 340; Indels 72; Gaps 18;

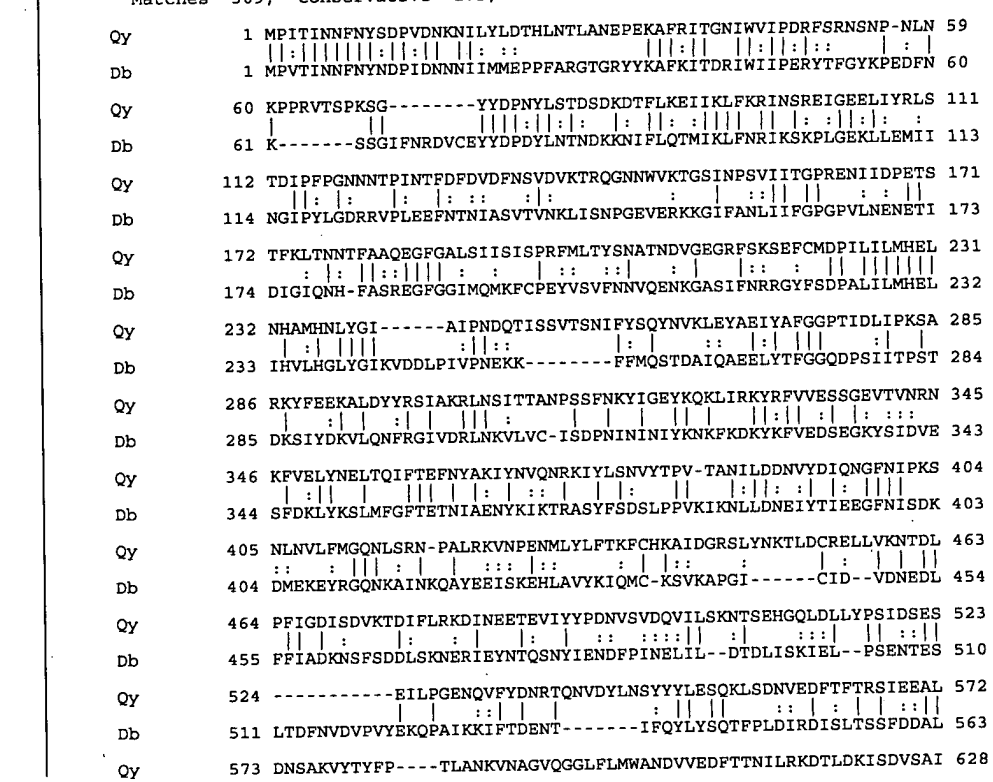
```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
```



RESULT 8
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-29

Query Match 20.2%; Score 1367; DB 4; Length 858;
Best Local Similarity 34.5%; Pred. No. 7.2e-79;
Matches 309; Conservative 175; Mismatches 340; Indels 72; Gaps 18;



```
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688

Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIK 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748

Db        681 TIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808

Db        741 YSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864

Db        801 TLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKY 856


RESULT 9
US-09-255-829-8
; Sequence 8, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-8

  Query Match           18.2%;  Score 1232.5;  DB 4;  Length 871;
  Best Local Similarity   33.7%;  Pred. No. 2.7e-70;
  Matches  303;  Conservative  163;  Mismatches  367;  Indels   67;  Gaps   23;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58

Db          1 MPFVNKQFNYKDPVNGVDIAYIKIP-NAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118

Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173

Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225

Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285

Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339

Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399

Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451

Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505

Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565

Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624

Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684

Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744

Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804

Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864

Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869


RESULT 10
US-09-255-829-2
; Sequence 2, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
```

```
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/255,829
;       FILING DATE:  23-FEB-1999
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB97/02273
;       FILING DATE:  22-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/782,893
;       FILING DATE:  27-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  ESMOND, ROBERT W.
;       REGISTRATION NUMBER:  32,893
;       REFERENCE/DOCKET NUMBER:  1581.0130002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-2600
;       TELEFAX:  202-371-2540
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  871 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-255-829-2

  Query Match             18.1%;  Score 1224.5;  DB 4;  Length 871;
  Best Local Similarity   33.6%;  Pred. No. 8.6e-70;
  Matches  302;  Conservative  163;  Mismatches  368;  Indels   67;  Gaps   23;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
              |     ||| |||:  :| |:    |    :| |||:|    |||||:|  |:     :|
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
              | ||        |||  |||||::|| :|| : |||:|| | ::|  |:  :   ||| |
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
                  :| | :   :| | ::|    |:   ::  :|  ::| ||  :||  |  :|
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
                 || |       |:|:   |  || |   :  :    ||  :| |:|:      || :
Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
              | ||| || | |||||| |    :  | :| :|    ::: : |:  |||      |
Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
                 |    : :: ||  ||        |||:          :| : : : ||    ::||:
Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
              :|:: || :||  ||:|:|| |: | : | ||| ||:          ||:     | | :||
Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
              |:  :||   | |||     |    |:      |: | |  | : |     :||   ||| |
Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505
              :   : | | || |      :   |:   ::|  :| :    :|:|:|   |  |:    |
Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
                |:  :: |   |  : |::|   : | : :   :|    ::|| :|:          |

Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
                |: ||| | ::|||:| :     |||   :  :|| |   :| |||       | |||:|
Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
              :: |||||||||| | : : :|  |   :|  |||  ||  || || | : |  :  :
Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

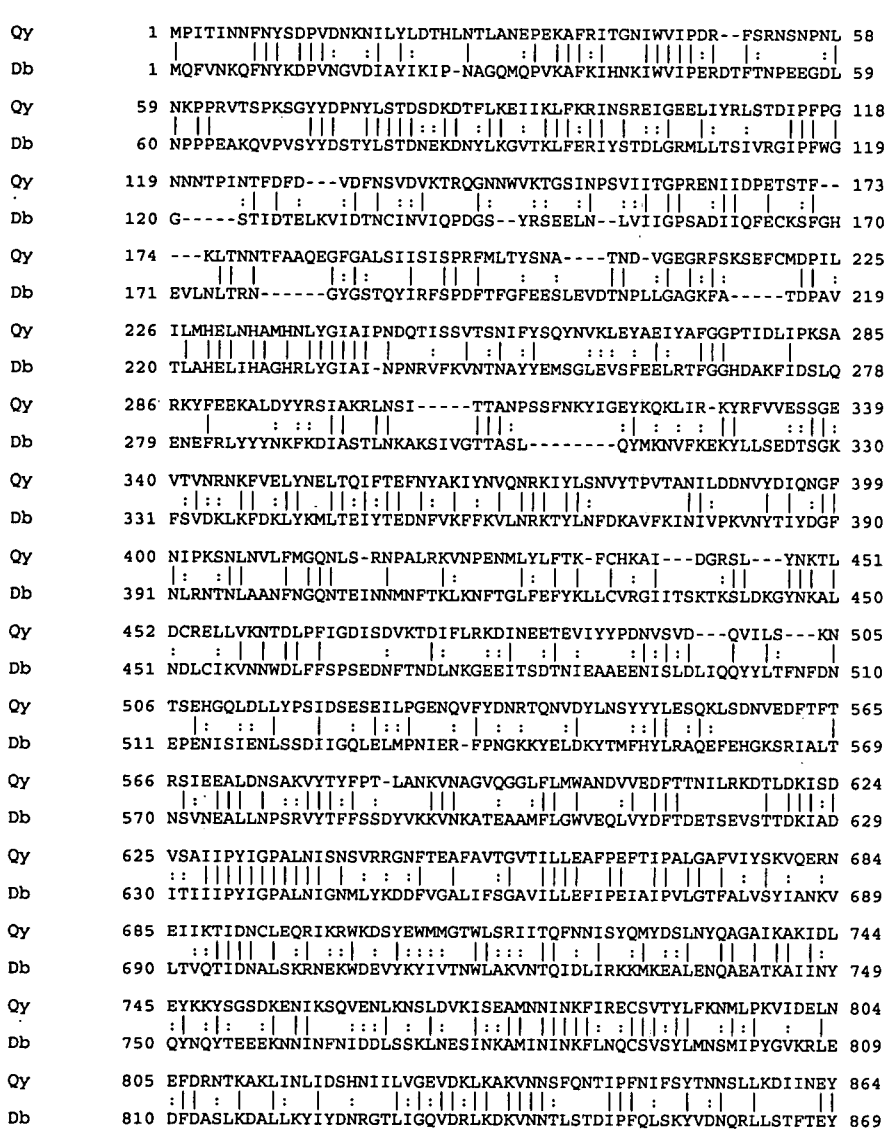
Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
                ::|||| | :| ::| : |::::  ||::: || : |  :| ::|  || | || |:
Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
              :| :|:  :| ||   :::| : |:  |::|| |||||: :|||:||  :|:|  :   |
Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864
              :|| : |  |:  |  :   |:|:||:|| ||||:     ||| :  | :|  |     ||
Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869


RESULT 11
US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-26

  Query Match             18.1%;  Score 1224.5;  DB 4;  Length 871;
  Best Local Similarity   33.6%;  Pred. No. 8.6e-70;
  Matches  302;  Conservative  163;  Mismatches  368;  Indels   67;  Gaps   23;
```



RESULT 12
US-09-255-829-16
; Sequence 16, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-16

Query Match 18.1%; Score 1224; DB 4; Length 907;
Best Local Similarity 32.7%; Pred. No. 9.8e-70;
Matches 308; Conservative 171; Mismatches 382; Indels 82; Gaps 26;

```
Qy        502 LS---KNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDN 558
Db        511 LTFNFDNEPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHG 569

Qy        559 VEDFTFTRSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKD 617
Db        570 KSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVS 629

Qy        618 TLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIY 677
Db        630 TTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALV 689

Qy        678 SKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGA 737
Db        690 SYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEA 749

Qy        738 IKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLP 797
Db        750 TKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIP 809

Qy        798 KVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLL 857
Db        810 YGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRL 869

Qy        858 KDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
Db        870 LSTFTEYIKSRPQSKV------KRQIF--SGYQSDIDTHNRIK 904
```

```
RESULT 13
US-09-255-829-10
; Sequence 10, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  875 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-09-255-829-10

  Query Match           18.1%;  Score 1220.5;  DB 4;  Length 875;
  Best Local Similarity  33.5%;  Pred. No. 1.6e-69;
  Matches  303;  Conservative  164;  Mismatches  366;  Indels   71;  Gaps   25;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
Db          1 MQFVNKQFNYKDPVNGVDIAYIKIP-NAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKSA 450

Qy        452 D--CRELLVK--NTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS- 503
Db        451 DGALNDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTF 510

Qy        504 --KNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVED 561
Db        511 NFDNEPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSR 569

Qy        562 FTFTRSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLD 620
Db        570 IALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTD 629

Qy        621 KISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKV 680
Db        630 KIADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYI 689

Qy        681 QERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKA 740
Db        690 ANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKA 749

Qy        741 KIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVI 800
Db        750 IINYQYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGV 809

Qy        801 DELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDI 860
Db        810 KRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLST 869

Qy        861 INEY 864
Db        870 FTEY 873
```

```
RESULT 14
```

```
US-09-255-829-6
; Sequence 6, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  873 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-6

  Query Match             18.0%;  Score 1216.5;  DB 4;  Length 873;
  Best Local Similarity   33.4%;  Pred. No. 2.8e-69;
  Matches  301;  Conservative  163;  Mismatches  369;  Indels   67;  Gaps   23;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
              |      ||| |||:  :| |:         :| |||:|   |||||:|  |:     :|
Db          3 MEFVNKQFNYKDPVNGVDIAYIKIP-KYGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 61

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
              | ||        |||  |||||::|| :|| : |||:|| | ::|  |:  :   ||| |
Db         62 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 121

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
                  :| | :   :| | ::|    |:   ::  :|  ::| ||  :||  |  :|
Db        122 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 172

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
                 || |      |:|:   |  || |   :  :    ||  :| |:|:      || :
Db        173 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 221

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
               | ||| || | |||||| |   :  | :| :|    ::: : |:  |||     |
Db        222 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 280

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
                 |     : :: ||  ||       |||:         :| : : : ||    ::||:
Db        281 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 332

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
              :|:: || :||  ||:|:|| |: | : | ||| ||:         ||:    | | :||
Db        333 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 392

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
              |:  :||   | |||     |    |:      |: | |  | : |     :||   ||| |
Db        393 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 452

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505
              :  : | |_|| |      :   |:  ::|  :| :    :|:|:|   |  |:    |
Db        453 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 512

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
                |:  :: |   |  : |::|   : | : :   :|    ::|| :|:          |
Db        513 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 571

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
              |: ||| | ::|||:| :    |||   :  :|| |   :| |||        | |||:|
Db        572 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 631

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
              :: |||||||||| | : : :|  |   :|  ||||  ||  || || | : | :  :
Db        632 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 691

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
              ::|||| | :| ::| : |::::  ||::: || : |  :| ::|  || | || |:
Db        692 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 751

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
              :| :|:  :| ||   :::| : |:  |::|| ||||: :|||:||  :|:|  :  |
Db        752 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 811

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864
              :|| : |  |:  |  :   |:|:||:|| ||||:    ||| :  | :|  |      ||
Db        812 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 871


RESULT 15
US-09-255-829-4
; Sequence 4, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-4

Query Match 18.0%; Score 1216.5; DB 4; Length 894;
Best Local Similarity 33.4%; Pred. No. 2.9e-69;
Matches 301; Conservative 163; Mismatches 369; Indels 67; Gaps 23;

```
Qy    1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
Db   24 MEFVNKQFNYKDPVNGVDIAYIKIP-KYGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 82

Qy   59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db   83 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 142

Qy  119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
Db  143 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 193

Qy  174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
Db  194 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 242

Qy  226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db  243 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 301

Qy  286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339
Db  302 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 353

Qy  340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399
Db  354 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 413

Qy  400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451
Db  414 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 473

Qy  452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505
Db  474 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 533

Qy  506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565
Db  534 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 592

Qy  566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
Db  593 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 652

Qy  625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684
Db  653 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 712

Qy  685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744
Db  713 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 772

Qy  745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804
Db  773 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 832

Qy  805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864
Db  833 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 892
```

Search completed: September 10, 2004, 03:47:08
Job time : 31.4018 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 249.897 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 6759
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6759	100.0	1291	12	US-10-272-898-60	Sequence 60, Appl
2	6759	100.0	1291	15	US-10-354-774-60	Sequence 60, Appl
3	6759	100.0	1291	15	US-10-271-012-60	Sequence 60, Appl
4	6759	100.0	1291	15	US-10-452-024-135	Sequence 135, App
5	6759	100.0	1291	15	US-10-205-516-6	Sequence 6, Appli
6	6759	100.0	1291	16	US-10-729-122-60	Sequence 60, Appl
7	6759	100.0	1291	16	US-10-729-039-60	Sequence 60, Appl
8	6753	99.9	1291	15	US-10-452-024-3	Sequence 3, Appli
9	6753	99.9	1291	15	US-10-452-024-136	Sequence 136, App
10	6747	99.8	1301	15	US-10-205-516-20	Sequence 20, Appl
11	5072.5	75.0	1280	15	US-10-452-024-139	Sequence 139, App
12	5049.5	74.7	1280	15	US-10-452-024-140	Sequence 140, App
13	4508.5	66.7	950	14	US-10-241-596-118	Sequence 118, App
14	4503.5	66.6	944	14	US-10-241-596-122	Sequence 122, App
15	4503	66.6	902	14	US-10-241-596-132	Sequence 132, App
16	4503	66.6	912	14	US-10-241-596-116	Sequence 116, App
17	4503	66.6	914	14	US-10-241-596-120	Sequence 120, App
18	4500	66.6	876	14	US-10-241-596-128	Sequence 128, App
19	4500	66.6	877	14	US-10-241-596-126	Sequence 126, App
20	4500	66.6	877	14	US-10-241-596-130	Sequence 130, App
21	4500	66.6	881	14	US-10-241-596-124	Sequence 124, App
22	4309	63.8	1285	15	US-10-452-024-166	Sequence 166, App
23	4307	63.7	1285	15	US-10-452-024-165	Sequence 165, App
24	4201.5	62.2	811	13	US-10-011-588-29	Sequence 29, Appl
25	3346.5	49.5	1286	15	US-10-205-516-22	Sequence 22, Appl
26	3343.5	49.5	1276	12	US-10-272-898-66	Sequence 66, Appl
27	3343.5	49.5	1276	15	US-10-354-774-66	Sequence 66, Appl
28	3343.5	49.5	1276	15	US-10-271-012-66	Sequence 66, Appl
29	3343.5	49.5	1276	15	US-10-452-024-164	Sequence 164, App
30	3343.5	49.5	1276	16	US-10-729-122-66	Sequence 66, Appl
31	3343.5	49.5	1276	16	US-10-729-039-66	Sequence 66, Appl
32	3338.5	49.4	1275	15	US-10-452-024-167	Sequence 167, App
33	3338.5	49.4	1276	15	US-10-452-024-4	Sequence 4, Appli
34	3338.5	49.4	1276	15	US-10-452-024-168	Sequence 168, App
35	3338.5	49.4	1276	15	US-10-205-516-8	Sequence 8, Appli
36	2891	42.8	552	15	US-10-452-024-141	Sequence 141, App
37	2400.5	35.5	824	13	US-10-011-588-33	Sequence 33, Appl
38	2386	35.3	450	10	US-09-910-186A-10	Sequence 10, Appl
39	2330	34.5	444	13	US-10-011-588-9	Sequence 9, Appli
40	2317	34.3	462	12	US-10-272-898-62	Sequence 62, Appl
41	2317	34.3	462	15	US-10-354-774-62	Sequence 62, Appl
42	2317	34.3	462	15	US-10-271-012-62	Sequence 62, Appl
43	2317	34.3	462	16	US-10-729-122-62	Sequence 62, Appl
44	2317	34.3	462	16	US-10-729-039-62	Sequence 62, Appl
45	2287	33.8	436	13	US-10-011-588-27	Sequence 27, Appl

ALIGNMENTS

```
RESULT 1
US-10-272-898-60
; Sequence 60, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;       SEQUENCE CHARACTERISTICS:
```

```
;           LENGTH: 1291 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-272-898-60

  Query Match           100.0%;  Score 6759;  DB 12;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 2
US-10-354-774-60
; Sequence 60, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-354-774-60

  Query Match           100.0%;  Score 6759;  DB 15;  Length 1291;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 3
US-10-271-012-60
; Sequence 60, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-271-012-60

  Query Match           100.0%;  Score 6759;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
```

```
Db        61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
Qy       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
Db       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
Qy       181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
Db       181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
Qy       241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
Db       241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
Qy       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
Db       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
Qy       361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
Db       361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
Qy       421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
Db       421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
Qy       481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
Db       481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
Qy       541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
Db       541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
Qy       601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
Db       601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
Qy       661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
Db       661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
Qy       721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
Db       721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
Qy       781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
Db       781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
Qy       841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
Db       841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
Qy       901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
Db       901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
Qy       961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
Db       961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
Qy      1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
Db      1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
Qy      1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
Db      1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
Qy      1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
Db      1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
Qy      1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
Db      1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
Qy      1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db      1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 4
US-10-452-024-135
; Sequence 135, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 135
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum phage
US-10-452-024-135

Query Match 100.0%; Score 6759; DB 15; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
Db         1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
Qy        61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
Db        61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
Qy       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
Db       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
Qy       181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
Db       181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
Qy       241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
Db       241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
Qy       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
Db       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
Qy       361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
Db       361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
Qy       421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
Db       421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
Qy       481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
Db       481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
```

```
Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 5
US-10-205-516-6
; Sequence 6, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-6

  Query Match             100.0%;  Score 6759;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy          61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy         121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy         181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy         241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy         301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy         361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy         421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy         481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy         541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy         601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy         661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy         721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy         781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy         841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy         901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy         961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy        1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 6
US-10-729-122-60
; Sequence 60, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-122-60

  Query Match           100.0%;  Score 6759;  DB 16;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
             |||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 7
US-10-729-039-60
; Sequence 60, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-039-60

  Query Match             100.0%;  Score 6759;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||||
```

```
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 8
US-10-452-024-3
; Sequence 3, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-3

  Query Match             99.9%;  Score 6753;  DB 15;  Length 1291;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 9
US-10-452-024-136
; Sequence 136, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 136
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-136

  Query Match             99.9%;  Score 6753;  DB 15;  Length 1291;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 10
US-10-205-516-20
; Sequence 20, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 1301
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-20

Query Match 99.8%; Score 6747; DB 15; Length 1301;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNK----TLDCRELLVKNTDLPFIGDISDVKTDI 476
              ||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKIEGRTLDCRELLVKNTDLPFIGDISDVKTDI 480

Qy        477 FLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDN 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDN 540
```

```
Qy        537 RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGG 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGG 600

Qy        597 LFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGV 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGV 660

Qy        657 TILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRI 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRI 720

Qy        717 ITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAM 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAM 780

Qy        777 NNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKV 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKV 840

Qy        837 NNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEE 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEE 900

Qy        897 GDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGY 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGY 960

Qy        957 TIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNM 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNM 1020

Qy       1017 MGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIF 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 MGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIF 1080

Qy       1077 AKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTR 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTR 1140

Qy       1137 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 1200

Qy       1197 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 1260

Qy       1257 DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||||||||
Db       1261 DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1295
```

```
RESULT 11
US-10-452-024-139
; Sequence 139, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 139
;   LENGTH: 1280
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
```

US-10-452-024-139

Query Match 75.0%; Score 5072.5; DB 15; Length 1280;
Best Local Similarity 77.5%; Pred. No. 0;
Matches 1002; Conservative 94; Mismatches 170; Indels 27; Gaps 10;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||:|||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRIIGNIWVIPDRFSRDSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSEKDTFLKEIIKLFKRINSREIGEELIYRLATDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNNVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQRISSVTSNIFYSQYKVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              |||||||||||||||||||||||||||||||||||| |:|||| ||| ||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVAVDRNKFAELYKELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDIKTDIFLSK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              |||||||| |||||||||||||||||||||||||||| |: ||::|||||||||||||||
Db        481 DINEETEVIDYPDNVSVDQVILSKNTSEHGQLDLLYPIIEGESQVLPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              |||||||||||||||||||||||| ||||||||| ||||||| ||:||| |||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTTSIEEALDNSGKVYTYFPKLADKVNTGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||| ||||||||||||||||||||||||||||||||||||||||||||||:||||||| |||
Db        661 EAFQEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMIGTWLSRITTQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDIKISEAMNNIN 780
              |||||||||||||| ||| ||||||||||||||||||||||||||||:|||||||||||
Db        721 NNISYQMYDSLNYQADAIKDKIDLEYKKYSGSDKENIKSQVENLKNSLDIKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||:||  || :|||||||||||||||:|||||| ||
Db        781 KFIRECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              :|||||||||||||||||||||||:||||||||||:|| |||||||||||  |||||
Db        841 ENTIPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTII 959
              :| |:  |||| ||||:  |:||  | ||:|:::||: |:||||:|:| ::|    ||||
Db        901 VNTIYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTII 957

Qy        960 DSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMG 1018
              :|:| |||| : | :  : : |:       :|: | |  | :  || |||||||:|||:||
Db        958 NSIKQNSGWKLCIRNGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMG 1017
```

```
Qy       1019 NMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAK 1078
Db       1018 YMKLYINGELKQSERIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSK 1069

Qy       1079 ELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN 1138
Db       1070 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPD 1129

Qy       1139 NNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHST 1195
Db       1130 RSKLYTGNPITIKSVSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSK 1189

Qy       1196 EDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRL 1254
Db       1190 NCVYALKLQSNLGNYGIG-IFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF 1246

Qy       1255 GGDWYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287
Db       1247 ------ENAYTP-VAVTNYETKLLSTSSFWKFI 1272
```

RESULT 12
US-10-452-024-140
; Sequence 140, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 140
; LENGTH: 1280
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-140

Query Match 74.7%; Score 5049.5; DB 15; Length 1280;
Best Local Similarity 77.2%; Pred. No. 0;
Matches 998; Conservative 94; Mismatches 174; Indels 27; Gaps 10;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRIIGNIWVIPDRFSRDSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
Db         61 PPRVTSPKSGYYDPNYLSTDSEKDTFLKEIIKLFKRINSREIGEELIYRLATDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
Db        181 AAQEGFGALSIISISPRFMLTYSNATNNVGEGRFSKSEFCMDPILILMHELNHTMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
Db        241 IAIPNDQRISSVTSNIFYSQYKVKLEYAEIYAFGGPTIDLIPKSGRKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVAVDRNKFAELYKELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDIKTDIFLSK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
Db        481 DINVETEVIDYPDNVSVDQVILSKNTSEHGQLDLLYPIIEGESQVLPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTTSIEEALDNSGKVYTYFPKLADKVNTGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRENFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
Db        661 EAFQEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMIGTWLSRITTQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
Db        721 NNISYQMYDSLNYQADAIKDKIDLEYKKYSGSDKENIKSQVENLKNSLDIKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
Db        781 KFIRECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
Db        841 ENTIPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTII 959
Db        901 VNTIYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTII 957

Qy        960 DSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMG 1018
Db        958 NSIKQNSGWKLCIRNGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMG 1017

Qy       1019 NMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAK 1078
Db       1018 YMKLYINGELKQSERIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSK 1069

Qy       1079 ELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN 1138
Db       1070 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPD 1129

Qy       1139 NNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHST 1195
Db       1130 RSKLYTGNPITIKSVSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSK 1189

Qy       1196 EDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRL 1254
Db       1190 NCVYALKLQSNLGNYGIG-IFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF 1246

Qy       1255 GGDWYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287
Db       1247 ------ENAYTP-VAVTNYETKLLSTSSFWKFI 1272
```

RESULT 13
US-10-241-596-118
; Sequence 118, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003

```
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 118
;   LENGTH: 950
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-118

  Query Match             66.7%;  Score 4508.5;  DB 14;  Length 950;
  Best Local Similarity   94.0%;  Pred. No. 4.8e-314;
  Matches  880;  Conservative   11;  Mismatches   22;  Indels   23;  Gaps    5;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 75

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 135

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 195

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 255

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 315

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 375

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 435

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 495

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 555

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 615

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 675

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 735

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 795

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 855

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN--NINDS----KILSLQNRKNTLVDTSGYNAEVS 894
              |||||||||||||||||||||||||    ::|     |::  ||  :  :|  |    :
Db        856 QNTIPFNIFSYTNNSLLKDIINEYFNLQTLDDPAFLYKVVIFQN--SDPLDCGGLTDTLQ 913

Qy        895 EEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIV 930
              | | ||                ||    : |:  |::
Db        914 AETD-QL--------------EDEKSALQTEIANLL 934


RESULT 14
US-10-241-596-122
; Sequence 122, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 122
;   LENGTH: 944
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-122

  Query Match             66.6%;  Score 4503.5;  DB 14;  Length 944;
  Best Local Similarity   96.8%;  Pred. No. 1.1e-313;
  Matches  875;  Conservative    6;  Mismatches   16;  Indels    7;  Gaps    3;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 75

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 135

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 195

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 255

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 315
```

```
Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 375

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 435

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 495

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 555

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 615

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 675

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 735

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 795

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 855

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN--NINDS----KILSLQNRKNTLVDTSGYNAEVS 894
              |||||||||||||||||||||||    ::|     |::  ||  : |  |  .  ||
Db        856 QNTIPFNIFSYTNNSLLKDIINEYFNLQTLDDPAFLYKVVIFQN-SDPLDLTDTLQAETD 914

Qy        895 EEGD 898
              :  |
Db        915 QLED 918


RESULT 15
US-10-241-596-132
; Sequence 132, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
;  APPLICANT: Microbiological Research Authority
;  APPLICANT:  The Speywood Laboratory Limited
;  TITLE OF INVENTION: Recombinant Toxin Fragments
;  FILE REFERENCE: 1581.0130003
;  CURRENT APPLICATION NUMBER: US/10/241,596
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 09/255,829
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: US 09/242,689
;  PRIOR FILING DATE: 1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/GB97/02273
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: US 08/782,893
;  PRIOR FILING DATE: 1996-12-27
;  PRIOR APPLICATION NUMBER: GB 9625996.5
;  PRIOR FILING DATE: 1996-12-13
;  PRIOR APPLICATION NUMBER: GB 9617671.4
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 175
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 132
;   LENGTH: 902
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-241-596-132

  Query Match             66.6%;  Score 4503;  DB 14;  Length 902;
  Best Local Similarity   98.4%;  Pred. No. 1.1e-313;
  Matches  870;  Conservative    4;  Mismatches    4;  Indels    6;  Gaps    2;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 75

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 135

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 195

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 255

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 315

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 375

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 435

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 495

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 555

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 615

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 675

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 735

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 795

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 855

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFN--NINDS----KILSLQN 878
              |||||||||||||||||||||||    ::|     |::  ||
Db        856 QNTIPFNIFSYTNNSLLKDIINEYFNLQTLDDPAFLYKVVIFQN 899


Search completed: September 10, 2004, 04:58:14
Job time : 253.897 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 22.8074 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 6759
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6759	100.0	1291	2	S46431	botulinum neurotox
2	6759	100.0	1291	2	A49777	botulinum neurotox
3	4309	63.8	1285	2	S70582	botulinum neurotox
4	3338.5	49.4	1276	2	S11455	botulinum neurotox
5	2891	42.8	552	2	A49928	neurotoxin C1, BoN
6	1909	28.2	1291	2	I40631	non-proteolytic bo
7	1908	28.2	1297	2	S39791	neurotoxin - Clost
8	1895	28.0	1291	1	A48940	bontoxilysin (EC 3
9	1828.5	27.1	1315	1	BTCLTN	tentoxilysin (EC 3
10	1822	27.0	1252	2	S21178	botulinum neurotox
11	1811.5	26.8	1251	2	JH0256	botulinum neurotox
12	1804.5	26.7	1268	2	S33411	botulinum neurotox
13	1777	26.3	1296	1	BTCLAB	bontoxilysin (EC 3
14	1768	26.2	1274	2	I40813	neurotoxin type F
15	1722	25.5	1296	2	I40645	botulinum neurotox
16	671	9.9	1162	2	I40817	botulinum toxin no
17	667	9.9	1162	2	A47708	progenitor toxin n
18	654.5	9.7	366	2	S48110	neurotoxin type F
19	651	9.6	1165	2	I40644	botulinum neurotox
20	641	9.5	367	2	S48106	neurotoxin type E
21	640	9.5	1196	2	S46430	botulinum neurotox
22	639	9.5	1196	2	JQ1467	toxin, nontoxic co
23	635.5	9.4	1193	2	S68218	botulinum neurotox
24	624	9.2	369	2	S48109	neurotoxin type F
25	616.5	9.1	1193	2	JC4901	nontoxic-nonhemagg
26	342.5	5.1	1127	2	T28317	ORF MSV156 hypothe
27	317	4.7	2269	2	T28677	rhoptry protein -
28	317	4.7	4688	2	F82885	hypothetical prote
29	307.5	4.5	2401	2	T28676	rhoptry protein -
30	297	4.4	960	2	S72284	DNA-directed RNA p
31	294.5	4.4	1436	2	D71618	hypothetical prote
32	290.5	4.3	1365	2	T18419	hypothetical prote
33	289	4.3	1711	2	T18429	hypothetical prote
34	288.5	4.3	4550	2	T18440	hypothetical prote
35	288	4.3	2485	1	H71621	serine/threonine-s
36	287	4.2	2910	2	T28156	DNA-directed RNA p
37	284.5	4.2	1183	2	F90559	conserved hypothet
38	282.5	4.2	3394	2	T18501	hypothetical prote
39	281.5	4.2	5005	2	F82884	hypothetical prote
40	281	4.2	3973	2	B71612	hypothetical prote
41	278	4.1	2166	2	G70163	hypothetical prote
42	275	4.1	2136	2	A05037	hypothetical prote
43	271.5	4.0	1830	2	E82909	conserved hypothet
44	271	4.0	2894	2	C64474	hypothetical prote
45	270.5	4.0	3724	2	T18427	hypothetical prote

ALIGNMENTS

RESULT 1
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 100.0%; Score 6759; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 6.4e-286;
Matches 1291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
```

```
Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 2
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.;
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxir
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutinir
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whol
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 100.0%; Score 6759; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 6.4e-286;
Matches 1291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
```

```
Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
```

RESULT 3
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membr
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein

F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

Query Match 63.8%; Score 4309; DB 2; Length 1285;
Best Local Similarity 65.1%; Pred. No. 1.5e-179;
Matches 840; Conservative 157; Mismatches 280; Indels 14; Gaps 6;

```
Qy          5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
              : :|||||||:: :||||     | |    | ||| || |||||||:||| ::||:|:||||
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRP 64

Qy         65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124
              ||    ||||:|||||  |||||| |||||||| |:||::||   |    || |:::||
Db         65 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 124

Qy        125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184
              :||||     :: |:  :  :|  |  | |||:| ||  ||:|   |       :  : |
Db        125 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 184

Qy        185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244
              ||| |||: ::| |:||:|: |::       || |||||:: ||||| |::| |||| ||
Db        185 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 244

Qy        245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304
              :|: |    |  |:||    ::: |:| ||| :::||:  |    |||| :|: |||||
Db        245 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGSDVEIIPQIERLQLREKALGHYKDIAKRL 304

Qy        305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364
              |:|    |||::  | :||:    || |  :::|   || :||  ||::|| : :|  |:
Db        305 NNINKTIPSSWSSNIDKYKKIFSEKYNFDKDNTGNFLVNIDKFNSLYSDLTNVMSEVIYS 364

Qy        365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424
              |||:||  | |  | || |||||||:| | ||||:     |:   |||: |||||:|:
Db        365 SQYNVKNRTHYFSKHYLPVFANILDDNIYTIINGFNLTTKGFNIENSGQNIERNPALQKL 424

Qy        425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484
              : |::: |||| |      |  |   |  : |||  ||:: |   :  :||  : | :
Db        425 SSESVVDLFTKVCL-----RLTRNSRDDSTCIQVKNNTLPYVADKDSISQEIFESQIITD 479

Qy        485 ETEVIYYPDNVSVDQVIL-SKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDY 543
              || |  | || |:|: || :|  :    :| | |::: |   :||| :|||| |::|||
Db        480 ETNVENYSDNFSLDESILDAKVPTNPEAVDPLLPNVNMEPLNVPGEEEVFYDDITKDVDY 539

Qy        544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603
              ||||||||:||||:|||: | | |:||||  | |:||: |:|| ||| ||| |||| |||
Db        540 LNSYYYLEAQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 599

Qy        604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663
              :|||||||||::|||||||||||||||||||||||| ||  |||| :|||  ||  ||| |
Db        600 EVVEDFTTNIMKKDTLDKISDVSAIIPYIGPALNIGNSALRGNFKQAFATAGVAFLLEGF 659

Qy        664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723
              |||||||| |  || :||| :|||||:||||||:|||||||:||:  ||||| |||||:|
Db        660 PEFTIPALGVFTFYSSIQERKKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHI 719

Qy        724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783
              ||||||||:||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 SYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 779

Qy        784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNT 843
              |||||||||||||||||||:||  || :|||||||||||||||:|||||| ||:||
Db        780 RECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESFENT 839

Qy        844 IPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNP 903
              |||||||||||||||:||||||:||||||||||:|||:|||||||| ||:|||||||||||
Db        840 IPFNIFSYTNNSLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNP 899

Qy        904 IFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVK 963
              |||||||||||:||||:||||||||||:|||||||||||||||||||||||||||||||||
Db        900 IFPFDFKLGSSGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSNLPGYTIIDSVK 959

Qy        964 NNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIY 1023
              ||||||||||||||||||||:||| |||||||| || |||||||||:| |||||| ||
```

```
Db  /     960 NNSGWSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFVTITTNMMGNMMIY 1019

Qy       1024 INGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGK 1083

Db       1020 INGKLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDK 1079

Qy       1084 DINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFN 1143

Db       1080 DINILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFN 1139

Qy       1144 EGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAIG 1202

Db       1140 EGYKIIIKRIRGNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYKPSRNLGTDLVPLG 1194

Qy       1203 LREQTKD-INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGD-WYR 1260

Db       1195 ALDQPMDEIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLGDDYWFN 1254

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291

Db       1255 HEYLIPVIKIEHYASLLESTSTHWVFVPASE 1285


RESULT 4
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             49.4%;  Score 3338.5;  DB 2;  Length 1276;
  Best Local Similarity   52.9%;  Pred. No. 2e-137;
  Matches  682;  Conservative  200;  Mismatches  375;  Indels   33;  Gaps   13;

Qy           5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64

Db           5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRP 64

Qy          65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124

Db          65 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 124

Qy         125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184

Db         125 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 184

Qy         185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244

Db         185 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 244

Qy         245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304

Db         245 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGHYKDIAKRL 304

Qy         305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364

Db         305 NNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSEVVYS 364

Qy         365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424

Db         365 SQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPALQKL 424

Qy         425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484

Db         425 SSESVVDLFTKVCL-----RLTKNSRDDSTCIKVKNNRLPYVADKDSISQEIFENKIITD 479

Qy         485 ETEVIYYPDNVSVDQVILSKNTSEHGQL-DLLYPSIDSESEILPGENQVFYDNRTQNVDY 543

Db         480 ETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFYDDITKYVDY 539

Qy         544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603

Db         540 LNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 599

Qy         604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663

Db         600 EVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFATAGVAFLLEGF 659

Qy         664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723

Db         660 PEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHI 719

Qy         724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783

Db         720 NYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 779

Qy         784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNT 843

Db         780 RECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVDRLKAKVNESFENT 839

Qy         844 IPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNP 903

Db         840 MPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNT 899

Qy         904 IFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSV 962

Db         900 IYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSI 956

Qy         963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMK 1021

Db         957 EQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMK 1016

Qy        1022 IYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 1081

Db        1017 LYINGELKQSQKIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELS 1068

Qy        1082 GKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNND 1141

Db        1069 NEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSK 1128

Qy        1142 FNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDI 1198

Db        1129 LYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCV 1188

Qy        1199 YAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGD 1257

Db        1189 YALKLQSNLGNYGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS---- 1241

Qy        1258 WYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287

Db        1242 -FKNAYTPVAVT--NYETKLLSTSSFWKFI 1268


RESULT 5
A49928
neurotoxin C1, BoNT/C1 - Clostridium botulinum phage CN
C;Species: Clostridium botulinum phage CN
C;Date: 13-Sep-1994 #sequence_revision 25-Apr-1997 #text_change 28-Jul-2000
C;Accession: A49928
R;Hauser, D.; Gibert, M.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
J. Bacteriol. 175, 7260-7268, 1993
A;Title: Comparative analysis of C3 and botulinal neurotoxin genes and their environment
A;Reference number: A49928; MUID:94042900; PMID:8226673
A;Accession: A49928
A;Status: preliminary
```

A;Molecule type: DNA
A;Residues: 1-552 <HAU>
A;Note: sequence extracted from NCBI backbone (NCBIN:139803, NCBIP:139804)
A;Note: strain A028
C;Superfamily: tetanus toxin

Query Match 42.8%; Score 2891; DB 2; Length 552;
Best Local Similarity 100.0%; Pred. No. 1.7e-118;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLES 552
              ||||||||||||
Db        541 VDYLNSYYYLES 552
```

RESULT 6
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 28.2%; Score 1909; DB 2; Length 1291;
Best Local Similarity 33.9%; Pred. No. 2.3e-75;
Matches 463; Conservative 255; Mismatches 493; Indels 156; Gaps 38;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||:|||||||:||:|| ||: ::            |||:||  ||:||:|::    | : |
Db          1 MPVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
              |         ||        ||||:||:|:  |: | : :|||| || |: :||:|:  :
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFFQTLIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
              ||: |:   |:  |: ::   :|:         :   |  ::|| ||   : : ||
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
              : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
              | :| ||||      :||::          |: |    ::  |:| |||     :|  |
Db        233 IHVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDTIQAEELYTFGGQDPSIISPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
              |  :| |  :| |  ||| :     |  |  |  || |   ||:|| :| |: :::
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
              | :||  |   ||| | |: | :: |  | |:   ||   |:||: :| |: ||||
Db        344 SFNKLYKSLMLGFTEINIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
              |:   : ||| : |  |  ::: |::    : | |::  :      | :  | | :|
Db        404 NMGKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKVPGI------CID--VDNENL 454

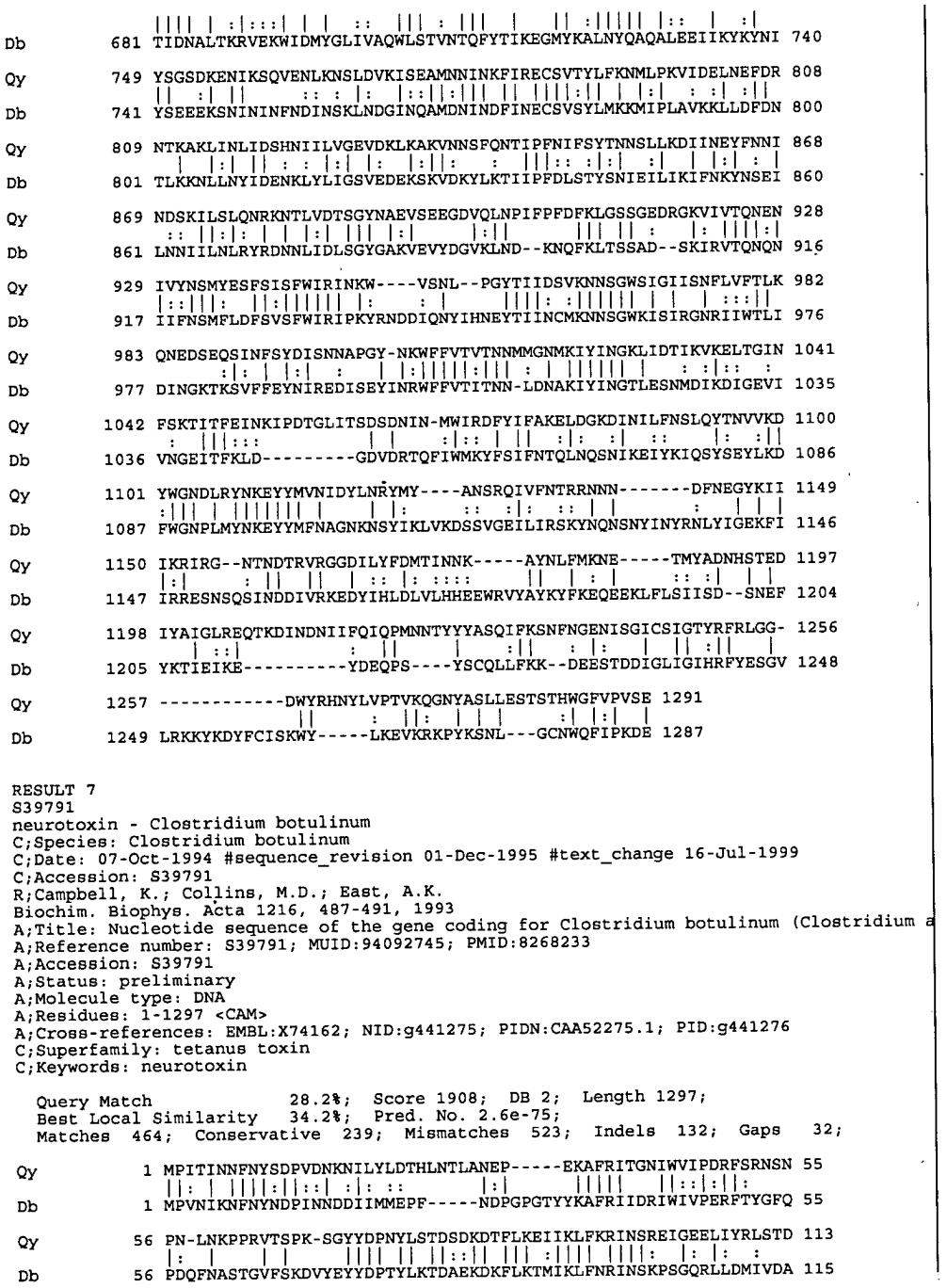
Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES 523
              || | :   |:  : :  |:  |  ::  ::::||  :|   :::|  || ::||
Db        455 FFIADKNSFSDDLSKNERVEYNTQNNYIGNDFPINELIL--DTDLISKIEL--PSENTES 510

Qy        524 ----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
                       |  |   :|| |  |        : || ||    |: | : | | ::||
Db        511 LTDFNVDVPVYEKQPAIKKVFTDENT-------IFQYLYSQTFPLNIRDISLTSSFDDAL 563

Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
              |:|||::|        ||||   |: |||  |   :|:||      :  |:|||:|:| |
Db        564 LVSSKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
              :|||| |||: :   :|||  || : | :||||  ||  || :|  |:: | :  :|:|||
Db        621 VPYIGLALNVGDETAKGNFESAFEIAGSSILLEFIPELLIPVVGVFLLESYIDNKNKIIK 680

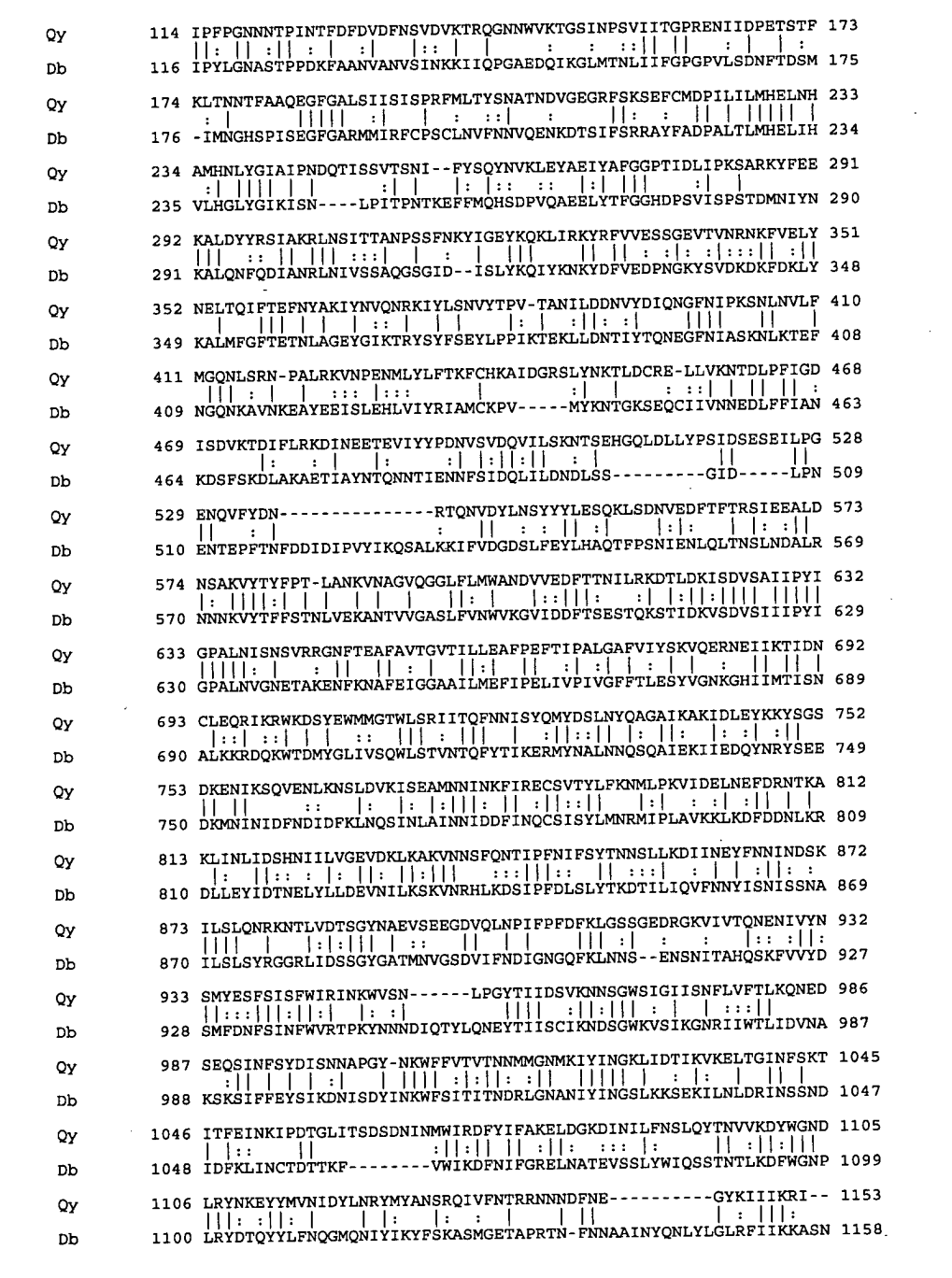
Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
```



RESULT 7
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 28.2%; Score 1908; DB 2; Length 1297;
Best Local Similarity 34.2%; Pred. No. 2.6e-75;
Matches 464; Conservative 239; Mismatches 523; Indels 132; Gaps 32;





```
Qy       1154 -RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDIND 1212
              |   ||  || || :| ::                  || | |  | : :   :|:|
Db       1159 SRNINNDNIVREGDYIYLNI----------------DNISDES-YRVYVLVNSKEIQT 1199

Qy       1213 NIIFQIQPMNN--TYYYASQI----FKSNFNGENIS-------GICSIGTYRFRLGGDWY 1259
               :   : |:|:  |:|   ||     |: :| : :        |:  || :    | |
Db       1200 QLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWD 1257

Qy       1260 RH-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1291
              : ||    : |     : |: :      :| |:|| |
Db       1258 TYDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 1293
```

RESULT 8
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 28.0%; Score 1895; DB 1; Length 1291;
Best Local Similarity 33.9%; Pred. No. 9.4e-75;
Matches 462; Conservative 248; Mismatches 504; Indels 150; Gaps 39;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||:|||||||:||:||  ||: ::           |||:||  ||:||:|::    | : |
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
              |         ||        ||||:||:|:  |: ||: :|||| || |: :||:|:  :
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
               ||: |:   |:  |: ::   :|:          :   |  ::|| ||   : : ||
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
               : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
               | :| ||||       :||::          |: |    ::   |:| |||   :|  |
Db        233 IHVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIITPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
               |  :| |  :| |  ||| :     |  |  |  || |   ||:|| :| |: :::
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
               | :||  |   ||| | |: | :: |  | |:   ||   |:||: :| |: ||||
Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
              ::   : ||| : |  |  ::: |::    : | |::  :      | :  | | ||
Db        404 DMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES 523
               || | :    |:  : |   |:  |  ::  ::::||  :|    :::|  || ::||
Db        455 FFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELIL--DTDLISKIEL--PSENTES 510

Qy        524 ----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
                         |  |   ::| |  |          : || ||    :: | : | | ::||
Db        511 LTDFNVDVPVYEKQPAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDAL 563
```

```
Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIK 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
Db        681 TIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
Db        741 YSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
Db        801 TLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
Db        861 LNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA--NSKIRVTQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
Db        917 IIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLI 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
Db        977 DINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKD 1100
Db       1036 ANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKD 1086

Qy       1101 YWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN-------DFNEGYKII 1149
Db       1087 FWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFI 1146

Qy       1150 IKRIRGNT---NDTRVRGGDILYFD-MTINNK----AYNLFMKNETMYADNHSTEDIYAI 1201
Db       1147 IRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE---------EKLFLA 1196

Qy       1202 GLREQTKDINDNIIFQIQPMNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRLGG---- 1256
Db       1197 PISDSDEFYN---TIQIKEYDEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYESGIVFE 1251

Qy       1257 ---------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db       1252 EYKDYFCISKWY-----LKEVKRKPYNLKL---GCNWQFIPKDE 1287
```

RESULT 9
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Es
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal h
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (frag
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglios
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized,
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of syn
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapto
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match   27.1%; Score 1828.5; DB 1; Length 1315;
  Best Local Similarity 32.6%; Pred. No. 7.4e-72;
  Matches 448; Conservative 253; Mismatches 526; Indels 147; Gaps   33;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
Db         61 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 120

Qy        120 NNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKL---T 176
Db        121 SYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDN 180

Qy        177 NNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMH 236
```

```
Db        181 KNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLH 240

Qy        237 NLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDY 296

Db        241 GLYGMQVSSHEIIPS--KQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLND 298

Qy        297 YRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQ 356

Db        299 YKAIANKLSQVTSCNDPNID--IDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMY 356

Qy        357 IFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLFMGQNL 415

Db        357 GFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNM 416

Qy        416 SRNP-ALRKVNPENMLYLFTKFCHKAID----GRSLYNKTLDCREL------LVKNTDLP 464

Db        417 RVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLT 476

Qy        465 FIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESE 524

Db        477 FIAEKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNL----QSKITLPN-DRTTP 531

Qy        525 ILPG-------ENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAK 577

Db        532 VTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTK 591

Qy        578 VYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637

Db        592 IYSYFPSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALN 651

Qy        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQR 697

Db        652 IVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKR 711

Qy        698 IKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENI 757

Db        712 YEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQI 771

Qy        758 KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINL 817

Db        772 ADEINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQY 831

Qy        818 IDSHN-IILVGEVDKLKAKVNNSFQNTIPF----NIFSYTNNSLLKDIINEYFNNINDSK 872

Db        832 IKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVI------LKKST 885

Qy        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGKVIVTQ 925

Db        886 ILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSEVIVHK 937

Qy        926 NENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWSIGIIS 974

Db        938 AMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKG 997

Qy        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032

Db        998 NNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSA 1056

Qy       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092

Db       1057 EITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEKLYTSY 1108

Qy       1093 QYTNVVKDYWGNDLRYNKEYYMVNI------------DYL---NRYMYANSRQIVFNTR 1136

Db       1109 LSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRR 1168

Qy       1137 RNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLFMKNET 1187

Db       1169 LYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDR 1223

Qy       1188 MYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI 1247

Db       1224 ILRVGYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL--KLYDDKNAS-LGLV 1268

Qy       1248 GTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291

Db       1269 GTHNGQIGNDPNRDILIASNWY-FNHLKDKI-----------LGCDWYFVPTDE 1310


RESULT 10
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match             27.0%;  Score 1822;  DB 2;  Length 1252;
  Best Local Similarity   33.8%;  Pred. No. 1.3e-71;
  Matches  451;  Conservative  256;  Mismatches  487;  Indels  142;  Gaps   43;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
```

```
Db          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Qy         61 PPRVTSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFP 117

Db         57 PP--TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114

Qy        118 GNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTN 177

Db        115 GNDNTPDNQFHIG-DASAVEIKFSNGSQDI----LLPNVIIMGAEPDLFETNSSNISLRN 169

Qy        178 NTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHN 237

Db        170 NYMPSNHGFGSIAIVTFSPE----YSFRFND-----NSMNEFIQDPALTLMHELIHSLHG 220

Qy        238 LYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDY 296

Db        221 LYGAKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNLLAD 278

Qy        297 YRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQ 356

Db        279 YKKIASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS 333

Qy        357 IFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLS 416

Db        334 -FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNAN 390

Qy        417 RNP-ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTD 475

Db        391 LNPRIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASENSYNDD 446

Qy        476 IFLRKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDSESEILP 527

Db        447 -----NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSDEKLNLT 491

Qy        528 GENQVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKV 578

Db        492 IQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKI 551

Qy        579 YTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637

Db        552 YTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALN 611

Qy        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIKTIDNCL 694

Db        612 IGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNAL 671

Qy        695 EQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDK 754

Db        672 KERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEK 731

Qy        755 ENI--KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKA 812

Db        732 NELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKT 791

Qy        813 KLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSK 872

Db        792 YLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSS 851

Qy        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYN 932

Db        852 VLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYIIYD 909

Qy        933 SMYESPSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDS 987

Db        910 NKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGI 969

Qy        988 EQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTI 1046

Db        970 NQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNI 1029

Qy       1047 TFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDL 1106

Db       1030 LFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1107 RYNKEYYMVNIDYLNRYM--YANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RGNTNDT 1160

Db       1081 LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN 1140

Qy       1161 RVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDIND--NIIFQI 1218

Db       1141 LVRKNDQVYINFVASKT--HLF----PLYADTATTN-------KEKTIKISSSGNRFNQV 1187

Qy       1219 QPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYA 1274

Db       1188 VVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVASTWY-------------YT 1227

Qy       1275 SLLESTSTH---WGFV 1287

Db       1228 HMRDHTNSNGCFWNFI 1243
```

RESULT 11
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.;
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxin
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release a
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 26.8%; Score 1811.5; DB 2; Length 1251;
Best Local Similarity 33.1%; Pred. No. 3.8e-71;
Matches 444; Conservative 261; Mismatches 482; Indels 155; Gaps 43;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEP------EKAFRITGNIWVIPDRFSRNS 54

Db          1 MP-TINSFNYNDPVNNRTILYI---------KPGGCQQFYKSFNIMKNIWIIPERNVIGT 50

Qy         55 NPNLNKPPRVTSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111

Db         51 IPQDFLPP--TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELS 108

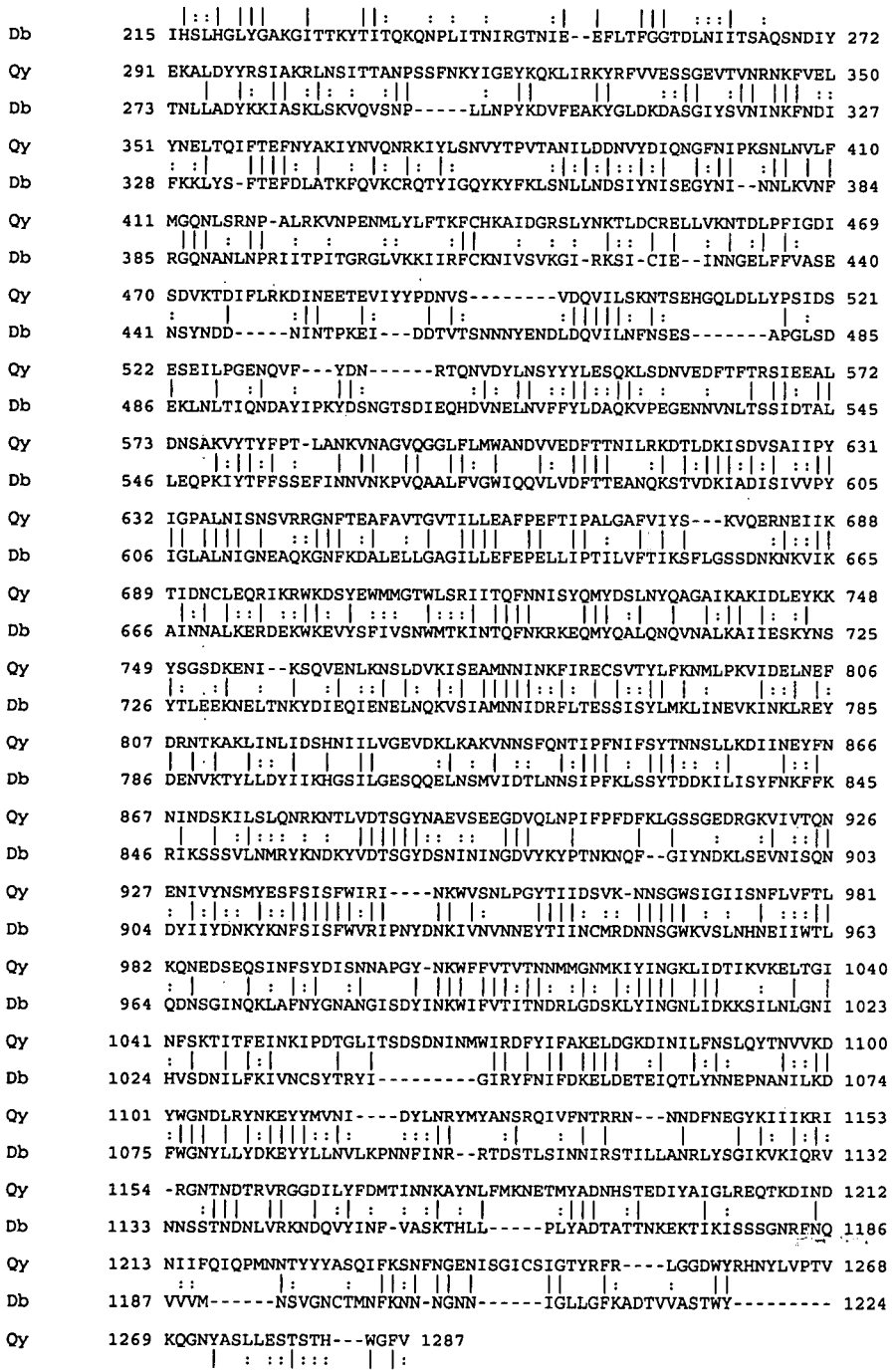
Qy        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171

Db        109 KANPYLGNDNTPDGDFIIN-DASAVPIQFSNGSQSI----LLPNVIIMGAEPDLFETNSS 163

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231

Db        164 NISLRNNYMPSNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHEL 214

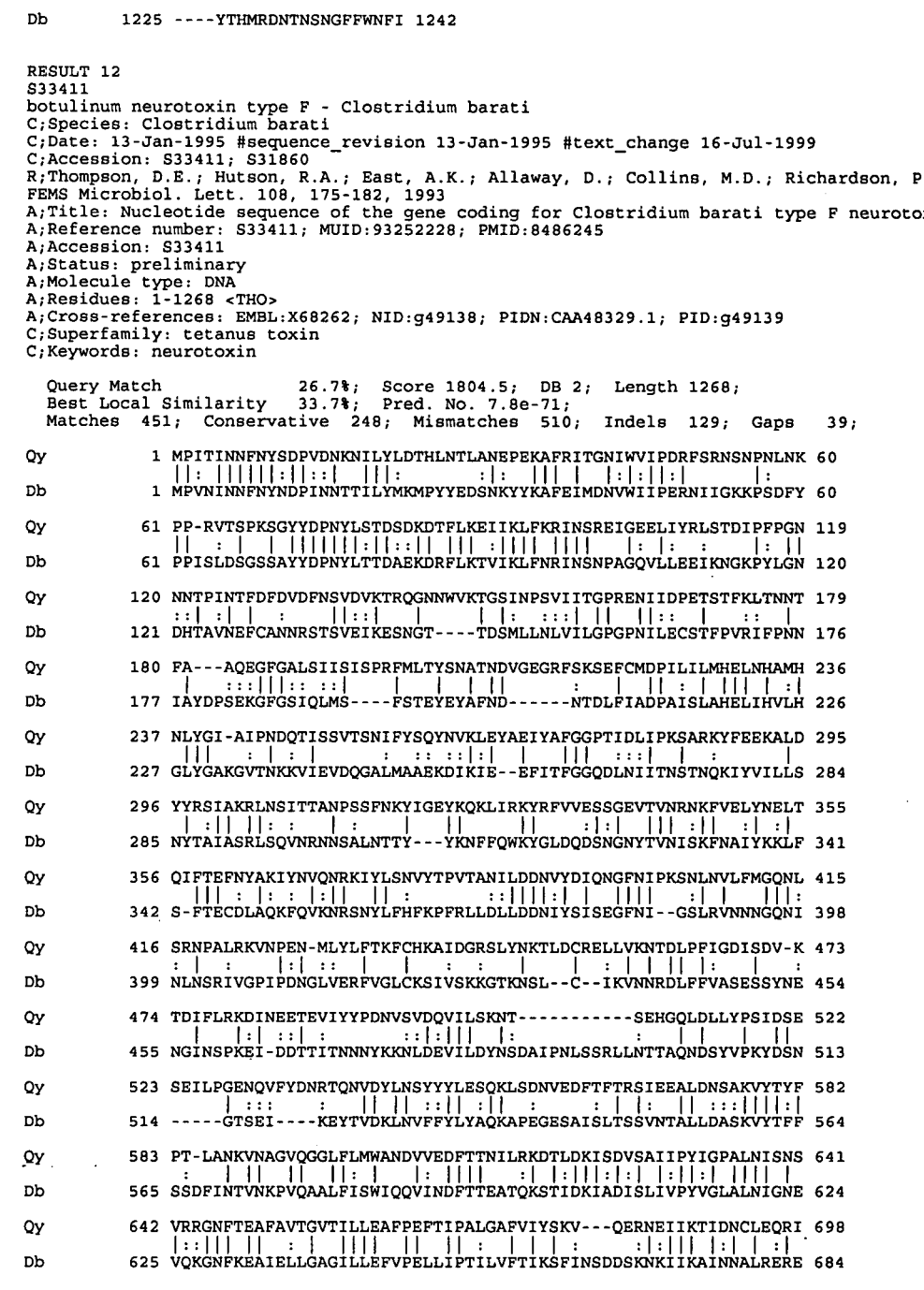
Qy        232 NHAMHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFE 290
```



Db 1225 ----YTHMRDNTNSNGFFWNFI 1242

RESULT 12
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 26.7%; Score 1804.5; DB 2; Length 1268;
Best Local Similarity 33.7%; Pred. No. 7.8e-71;
Matches 451; Conservative 248; Mismatches 510; Indels 129; Gaps 39;



```
Qy        699 KRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIK 758
Db        685 LKWKEVYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDGIKKIIEYKYNNYTLDEKNRLR 744

Qy        759 SQ--VENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLIN 816
Db        745 AEYNIYSIKEELNKKVSLAMQNIDRFLTESSISYLMKLINEAKINKLSEYDKRVNQYLLN 804

Qy        817 -LIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILS 875
Db        805 YILENSSTLGTSSVPELNNLVSNTLNNSIPFELSEYTNDKILIHILIRFYKRIIDSSILN 864

Qy        876 LQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMY 935
Db        865 MKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYSS--RLSEVNITQNNTIIYNSRY 922

Qy        936 ESFSISFWIRINKW--VSNLPG-YTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDSEQSI 991
Db        923 QNFSVSFWVRIPKYNNLKNLNNEYTIINCMRNNNSGWKISLNYNNIIWTLQDTTGNNQKL 982

Qy        992 NFSY----DISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTIT 1047
Db        983 VFNYTQMIDISDY---INKWTFVTITNNRLGHSKLYINGNLTDQKSILNLGNIHVDDNIL 1039

Qy       1048 FEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLR 1107
Db       1040 FKIVGCNDTRYV---------GIRYFKIFNMELDKTEIETLYHSEPDSTILKDFWGNYLL 1090

Qy       1108 YNKEYYMVNIDYLNRYMYANSRQIVFNTRR---------NNNDFNEGYKIIIKRIRGNTN 1158
Db       1091 YNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNIFSNARLYTGVEVIIRKV-GSTD 1149

Qy       1159 ----DTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLRE-QTKDINDN 1213
Db       1150 TSNTDNFVRKNDTVYINVVDGNSEYQL-------YADVSTSAVEKTIKLRRISNSNYNSN 1202

Qy       1214 IIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRL----GGDWYRHNYLVPTVK 1269
Db       1203 ---QMIIMDSIGDNCTMNFKTN-NGND------IGLLGFHLNNLVASSWYYKNIRNNTRN 1252

Qy       1270 QGNYASLLESTSTHWGFV 1287
Db       1253 NGCF----------WSFI 1260
```

RESULT 13
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-termi
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ext
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arour
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; N
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synape
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prote
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 26.3%; Score 1777; DB 1; Length 1296;
Best Local Similarity 32.9%; Pred. No. 1.2e-69;
Matches 444; Conservative 243; Mismatches 542; Indels 122; Gaps 34;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNL 58
Db          1 MPFVNKQFNYKDPVNGVDIAYIKIP-NAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL 59

Qy         59 NKPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPG 118
Db         60 NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWG 119

Qy        119 NNNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF-- 173
Db        120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGH 170

Qy        174 ---KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPIL 225
```

```
Db        171 EVLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAV 219

Qy        226 ILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285

Db        220 TLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Qy        286 RKYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGE 339

Db        279 ENEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGK 330

Qy        340 VTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGF 399

Db        331 FSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGF 390

Qy        400 NIPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTL 451

Db        391 NLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKAL 450

Qy        452 DCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KN 505

Db        451 NDLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDN 510

Qy        506 TSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFT 565

Db        511 EPENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALT 569

Qy        566 RSIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624

Db        570 NSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIAD 629

Qy        625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERN 684

Db        630 ITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKV 689

Qy        685 EIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDL 744

Db        690 LTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINY 749

Qy        745 EYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELN 804

Db        750 QYNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLE 809

Qy        805 EFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEY 864

Db        810 DFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEY 869

Qy        865 FNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKV 921

Db        870 IKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KI 924

Qy        922 IVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLV 978

Db        925 EVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEII 984

Qy        979 FTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKEL 1037

Db        985 WTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044

Qy       1038 TGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNV 1097

Db       1045 GNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 1096

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR-RNNNDFNE 1144

Db       1097 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYR 1156

Qy       1145 GYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLR 1204

Db       1157 GTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSA 1208

Qy       1205 EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYL 1264

Db       1209 LEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNI 1258

Qy       1265 VPTVKQGNYASLLESTS----THWGFVPVSE 1291

Db       1259 AKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289
```

RESULT 14
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 26.2%; Score 1768; DB 2; Length 1274;
Best Local Similarity 33.3%; Pred. No. 3e-69;
Matches 446; Conservative 251; Mismatches 519; Indels 124; Gaps 41;

```
Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59

Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTNPSDFD 60

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119

Db         61 PPASLKNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYLGN 120

Qy        120 NNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPR---------ENIIDPET 170

Db        121 DHTPIDEFSPVTRTTSVNIKL--STNVESSMLLNLLVLGAGPDIFESCCYPVRKLIDPDV 178

Qy        171 STFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSE-FCMDPILILMH 229

Db        179 -VYDPSN------YGFGSINIVTFSPEYEYTF----NDISGGHNSSTESFIADPAISLAH 227

Qy        230 ELNHAMHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKY 288

Db        228 ELIHALHGLYGARGVTYEETIEVKQAPLMIAEKPIRLE--EFLTFGGQDLNIITSAMKEK 285

Qy        289 FEEKALDYYRSIAKRLNSITTANPSSFNKY-IGEYKQKLIRKYRFVVESSGEVTVNRNKF 347

Db        286 IYNNLLANYEKIATRLSEVNSAPP----EYDINEYKDYFQWKYGLDKNADGSYTVNENKF 341

Qy        348 VELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLN 407

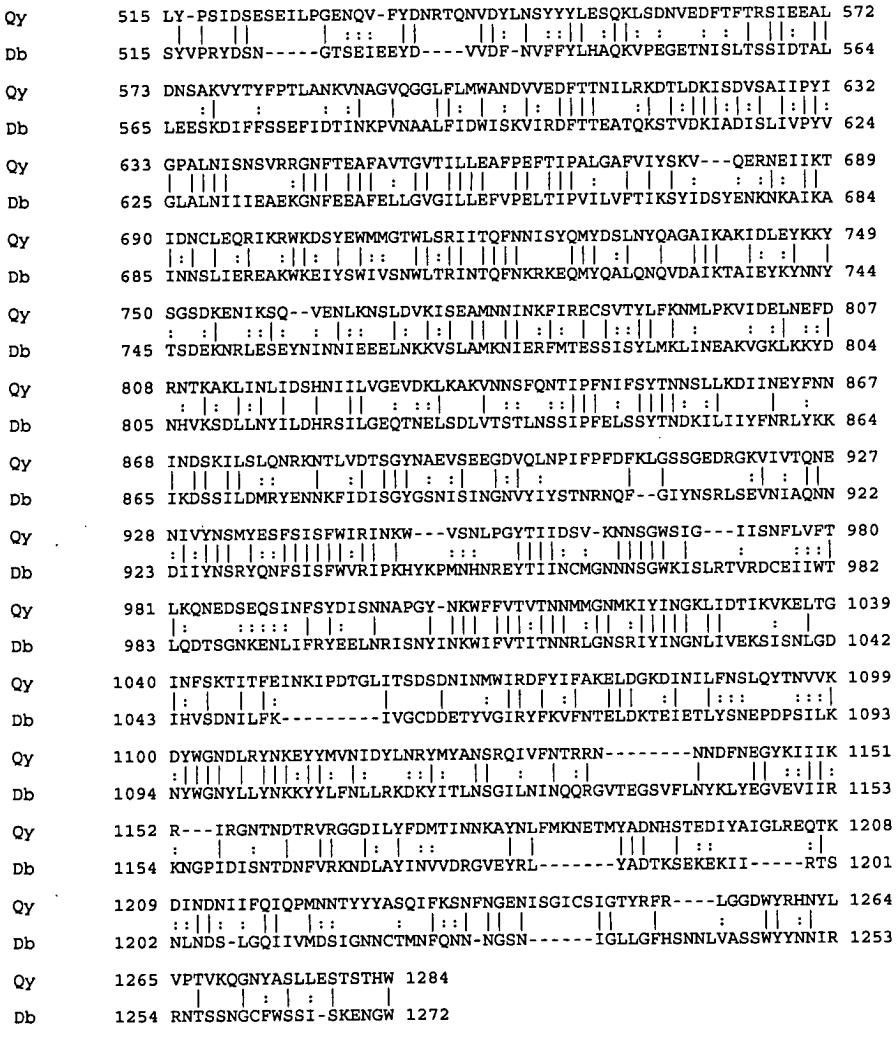
Db        342 NEIYKKLYS-FTESDLANKFKVKCRNTYFIKYEFLKVPNLLDDDIYTVSEGFNI--GNLA 398

Qy        408 VLFMGQNLSRNPALRKVNPE-NMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFI 466

Db        399 VNNRGQSIKLNPKIIDSIPDKGLVEKIVKFCKSVIPRKGTKAPPRLC--IRVNNSELFFV 456

Qy        467 GDISDV-KTDIFLRKDINEETEV-IYYPDNVSVDQVILSKNTSEHGQL----------DL 514

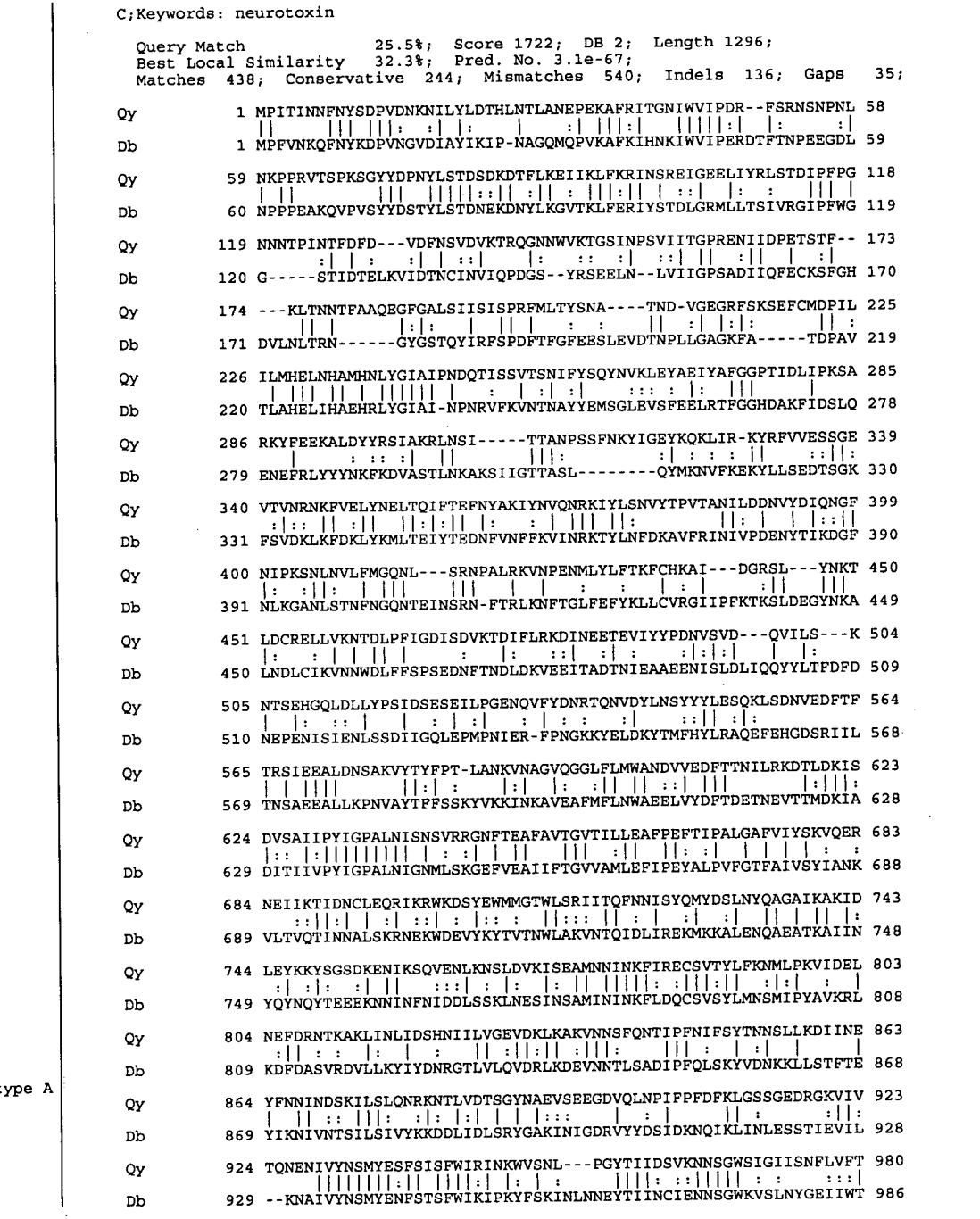
Db        457 ASESSYNENDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSQTIPQISNRTLNTLVQDN 514
```



RESULT 15
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin

C;Keywords: neurotoxin

Query Match 25.5%; Score 1722; DB 2; Length 1296;
Best Local Similarity 32.3%; Pred. No. 3.1e-67;
Matches 438; Conservative 244; Mismatches 540; Indels 136; Gaps 35;



```
Qy        981 LKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTG 1039
              |: |: : | : | |     |   | |:| |||:||| :   ||||||:|||   : |
Db        987 LQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQKPISNLGN 1046

Qy       1040 INFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVK 1099
              |: |  | |:::   |           : |: | :| |||: |:|  |::|   : ::|
Db       1047 IHASNKIMFKLDGCRD--------PRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1098

Qy       1100 DYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTRRNNNDFNEGY 1146
              |:||| |:|:| |||:|:   |:|:  |:             :|      |:`  ||
Db       1099 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNSTLYEGT 1158

Qy       1147 KIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQ 1206
              | |||:      |  ||  | :| :: : || | |         | | |   :  |    :
Db       1159 KFIIKKYASGNEDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSALE 1210

Qy       1207 TKDINDNIIFQI--------QPMNNTYYYASQIFKSNFNGENISGICSIGTYRF-----R 1253
                |: :  : |:        | : |          | |    | : |  || : :
Db       1211 IPDVGN--LSQVVVMKSKDDQGIRNK-------CKMNLQDNNGNDIGFIGFHLYDNIAKL 1261

Qy       1254 LGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              :  :||         :|   ||   :    | |:|| :
Db       1262 VASNWYN--------RQVGKAS--RTFGCSWEFIPVDD 1289
```

Search completed: September 10, 2004, 03:44:12
Job time : 27.8074 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 15.7519 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 6759
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6748	99.8	1290	1	BXC1_CLOBO	P18640 clostridium
2	3338.5	49.4	1276	1	BXD_CLOBO	P19321 clostridium
3	1896	28.1	1296	1	BXG_CLOBO	Q60393 clostridium
4	1890	28.0	1290	1	BXB_CLOBO	P10844 clostridium
5	1823.5	27.0	1314	1	TETX_CLOTE	P04958 clostridium
6	1810	26.8	1250	1	BXE_CLOBU	P30995 clostridium
7	1794	26.5	1250	1	BXE_CLOBO	Q00496 clostridium
8	1772	26.2	1295	1	BXA1_CLOBO	P10845 clostridium
9	1768	26.2	1274	1	BXF_CLOBO	P30996 clostridium
10	1717	25.4	1295	1	BXA2_CLOBO	Q45894 clostridium
11	671	9.9	1162	1	BXEN_CLOBU	Q06366 clostridium
12	665	9.8	1162	1	BXEN_CLOBO	P46082 clostridium
13	639	9.5	1196	1	BXCN_CLOBO	P46081 clostridium
14	275	4.1	2136	1	YCF2_MARPO	P09975 marchantia
15	271	4.0	2894	1	YD96_METJA	Q58791 methanococc
16	263	3.9	2710	1	TOXA_CLODI	P16154 clostridium
17	251.5	3.7	1169	1	EX5B_BORBU	O51578 borrelia bu
18	251	3.7	1630	1	MSP1_PLAFK	P04932 plasmodium
19	251	3.7	1639	1	MSP1_PLAFW	P04933 plasmodium
20	237	3.5	1076	1	RPOB_ASTLO	P27059 astasia lon
21	231	3.4	2339	1	RPC1_PLAFA	P27625 plasmodium
22	224	3.3	1102	1	RPOP_AGABT	P33539 agaricus bi
23	223.5	3.3	1024	1	RPOB_PLAFA	P21421 plasmodium
24	220	3.3	3135	1	S230_PLAFO	Q08372 plasmodium
25	219	3.2	2334	1	WAPA_BACSU	Q07833 bacillus su
26	219	3.2	2867	1	RBP2_PLAVB	Q00799 plasmodium
27	216	3.2	1701	1	MSP1_PLAFM	P08569 plasmodium
28	215	3.2	1682	1	MSP1_PLAF3	P19598 plasmodium
29	214	3.2	1176	1	SLAP_BACSH	P38537 bacillus sp
30	210.5	3.1	780	1	MUS2_BORBU	O51125 borrelia bu
31	209	3.1	1701	1	MSP1_PLAFF	P13819 plasmodium
32	209	3.1	1726	1	MSP1_PLAFC	P04934 plasmodium
33	209	3.1	1726	1	MSP1_PLAFP	P50495 plasmodium
34	208.5	3.1	2366	1	TOXB_CLODI	P18177 clostridium
35	208	3.1	1104	1	COLA_CLOPE	P43153 clostridium
36	208	3.1	1177	1	Y307_MYCGE	P47549 mycoplasma
37	208	3.1	2971	1	YCX9_CHLRE	Q32065 chlamydomon
38	206	3.0	4705	1	FAT2_DROME	Q9vw71 drosophila
39	205	3.0	986	1	EPIB_STAEP	P30195 staphylococ
40	202	3.0	1225	1	Y309_MYCGE	P47551 mycoplasma
41	201.5	3.0	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
42	200.5	3.0	1197	1	DPOM_PODAN	Q01529 podospora a
43	199.5	3.0	2368	1	ESR1_YEAST	P38111 saccharomyc
44	198.5	2.9	1803	1	YJL3_YEAST	P47024 saccharomyc
45	198	2.9	3097	1	CADN_DROME	O15943 drosophila

ALIGNMENTS

RESULT 1
BXC1_CLOBO
ID BXC1_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE (Bontoxilysin C1).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90370487; PubMed=2204031;
RA Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA Boquet P., Popoff M.R.;
RT "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL Nucleic Acids Res. 18:4924-4924(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=91024998; PubMed=2222445;
RA Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT "The complete nucleotide sequence of the gene coding for botulinum
RT type C1 toxin in the C-ST phage genome.";
RL Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN [3]
RP SEQUENCE OF 2-25.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=88153072; PubMed=2450068;
RA Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA Oguma K.;
RT "Establishment of a monoclonal antibody recognizing an antigenic site
RT common to Clostridium botulinum type B, C1, D, and E toxins and
RT tetanus toxin.";
RL Infect. Immun. 56:898-902(1988).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94038966; PubMed=7901002;
RA Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT cleaving HPC-1/syntaxin.";
RL EMBO J. 12:4821-4828(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             99.8%;  Score 6748;  DB 1;  Length 1290;
  Best Local Similarity   99.9%;  Pred. No. 2.7e-285;
  Matches 1289;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP 60

Qy         62 PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 121
              ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db         61 PRVTSPKSGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 120

Qy        122 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 180

Qy        182 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGI 240

Qy        242 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIA 300

Qy        302 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEF 360

Qy        362 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPAL 420

Qy        422 RKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKD 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKD 480

Qy        482 INEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNV 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 INEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNV 540

Qy        542 DYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMW 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMW 600

Qy        602 ANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLE 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLE 660

Qy        662 AFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFN 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFN 720

Qy        722 NISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINK 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINK 780

Qy        782 FIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQ 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 FIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQ 840

Qy        842 NTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQL 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQL 900

Qy        902 NPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDS 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 NPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDS 960

Qy        962 VKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMK 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 VKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMK 1020

Qy       1022 IYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 IYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 1080

Qy       1082 GKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNND 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNND 1140

Qy       1142 FNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 FNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1200

Qy       1202 GLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRH 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRH 1260

Qy       1262 NYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              |||||||||||||||||||||||||||||
Db       1261 NYLVPTVKQGNYASLLESTSTHWGFVPVSE 1290


RESULT 2
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
```

```
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match           49.4%;  Score 3338.5;  DB 1;  Length 1276;
  Best Local Similarity 52.9%;  Pred. No. 1.5e-137;
  Matches  682;  Conservative  200;  Mismatches  375;  Indels   33;  Gaps   13;

Qy          5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
              : :|||||||:: :||||    | |   | ||| || ||||||:||| ::||:|:||||
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRP 64

Qy         65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124
              ||    ||||:||||| |||||| |||||||| |:||::||  |    || |:::||
Db         65 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 124

Qy        125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184
              :|||     :: |:  :  :|  |  | |||:| ||  ||:|   |        :  : |
Db        125 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 184

Qy        185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244
              ||| |||: ::| |:||:|: |::       || |||||:: ||||| |::| |||| ||
Db        185 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 244

Qy        245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304
              :|: |    |  |:||   ::: |:| ||| :::||:  |    |||| :|: |||||
Db        245 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGHYKDIAKRL 304

Qy        305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364
              |:|    |||:   | :||:    || |  :::|   || :||  ||::|| : :|  |:
Db        305 NNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSEVVYS 364

Qy        365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424
              |||:||  | |  | || |||||||:| |::|||:     |:   |||: |||||:|:
Db        365 SQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPALQKL 424

Qy        425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484
              : |::: |||| |       |   |    |   : |||  ||:: |   :  :||  | | :
Db        425 SSESVVDLFTKVCL-----RLTKNSRDDSTCIKVKNNRLPYVADKDSISQEIFENKIITD 479

Qy        485 ETEVIYYPDNVSVDQVILSKNTSEHGQL-DLLYPSIDSESEILPGENQVFYDNRTQNVDY 543
              || |  | |  |:|: ||       : :: | | |::: |   ||||  ||||: |: |||
Db        480 ETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFYDDITKYVDY 539

Qy        544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603
              |||||||||||||:|||: | | |:|||  | |:||: |:|| ||| ||| |||| |||
Db        540 LNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 599
```

```
Qy        604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663
              :|||||||||::|||||||||| |||||||||| ||  |||| :|||  ||  ||| |
Db        600 EVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFATAGVAFLLEGF 659

Qy        664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723
              |||||||| |  || :||| :|||||:||||||:||||||:||:  ||||| ||||:|
Db        660 PEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHI 719

Qy        724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783
              :|||||||:||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        720 NYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 779

Qy        784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNT 843
              ||||||||||||||||||:||  || :|||||||||||||||:|||||| ||:||
Db        780 RECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVDRLKAKVNESFENT 839

Qy        844 IPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNP 903
              :|||||||||||||||||||:||||||||||:|| |||||||||||    :||||
Db        840 MPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNT 899

Qy        904 IFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSV 962
              |:  |||| |||:   |:||  | ||:|:::||: |:||||:|:|  ::|     ||||:|:
Db        900 IYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSI 956

Qy        963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMK 1021
              : |||| : | :  : : |:       :|: | |  | :  || ||||||:|||:|| ||
Db        957 EQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMK 1016

Qy       1022 IYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 1081
              :||||:|  : |:::|  :   ||| | |::  |         :|  :||||| ||:|||
Db       1017 LYINGELKQSQKIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELS 1068

Qy       1082 GKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNND 1141
               :||||::      ||:|||||| |::: |||::| :|::||:   |  :|     : :
Db       1069 NEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSK 1128

Qy       1142 FNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDI 1198
                |  | || :      :|:  || :   |  |:: | :    :|:||      |  :
Db       1129 LYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCV 1188

Qy       1199 YAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGD 1257
              ||: |:    :     || |: : :   | |||| |:|  ||   :  |   :||
Db       1189 YALKLQSNLGNYGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS---- 1241

Qy       1258 WYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287
               ::: |    |   || : | |||: | |:
Db       1242 -FKNAYTPVAVT--NYETKLLSTSSFWKFI 1268
```

```
RESULT 3
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             28.1%;  Score 1896;  DB 1;  Length 1296;
  Best Local Similarity   34.0%;  Pred. No. 5.1e-75;
  Matches  462;  Conservative  239;  Mismatches  524;  Indels  132;  Gaps   32;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEP-----EKAFRITGNIWVIPDRFSRNSNP 56
              |: |  |||:||::| :|: ::       |:|      |||||   ||::|:||:     |
Db          1 PVNIKXFNYNDPINNDDIIMMEPF-----NDPGPGTYYKAFRIIDRIWIVPERFTYGFQP 55

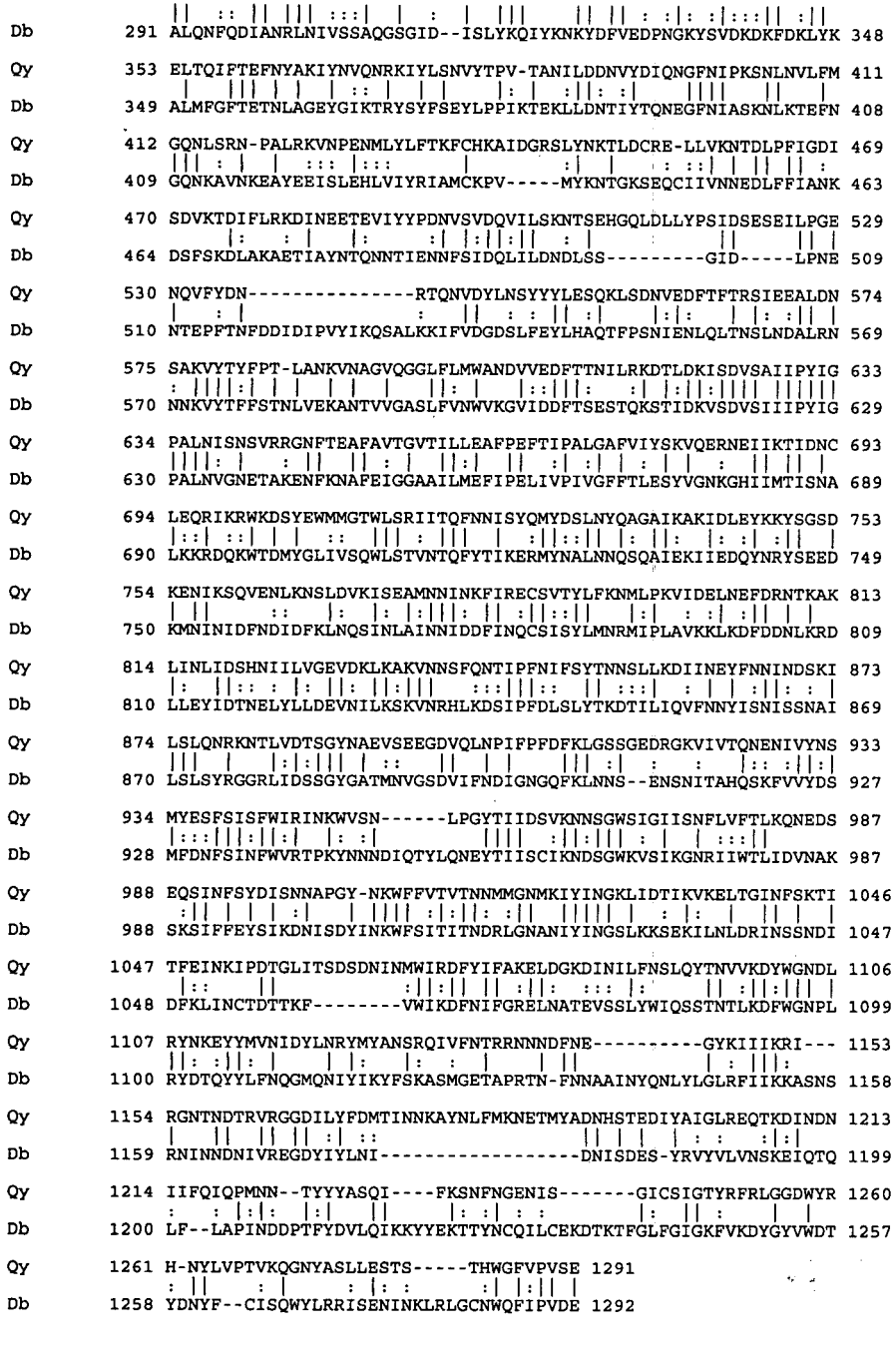
Qy         57 N-LNKPPRVTSPK-SGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDI 114
              :  |    | |     |||| || ||::|| ||| :|||| ||||:  |: |:  :   |
Db         56 DQFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAI 115

Qy        115 PFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFK 174
              |: || :|| : |  :|    |:: |  |      :   :  ::|| ||   : |  | :
Db        116 PYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSM- 174

Qy        175 LTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHA 234
              : |       |||||  :|   |  :  ::|   :     ||:  :  || | ||||| |
Db        175 IMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHV 234

Qy        235 MHNLYGIAIPNDQTISSVTSNI--FYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEK 292
              :| |||| | |       :| |   |: |::  ::  |:| |||    :|  |        |
Db        235 LHGLYGIKISN----LPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNK 290

Qy        293 ALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYN 352
```



RESULT 4

BXB_CLOBO
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence.";
RL Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.";
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT "Botulinum neurotoxins are zinc proteins.";
RL J. Biol. Chem. 267:23479-23483(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RL Nature 359:832-835(1992).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC SYNAPTOBREVIN-2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

```
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             28.0%;  Score 1890;  DB 1;  Length 1290;
  Best Local Similarity   33.8%;  Pred. No. 9.2e-75;
  Matches  461;  Conservative  248;  Mismatches  504;  Indels  150;  Gaps   39;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLNK 60
              |:|||||||:||:|| ||: ::          |||:||  ||:||:|::    | : ||
Db          1 PVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNK 60

Qy         61 PPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLST 112
                      ||        ||||:||:|:  |: ||: :|||| || |: :||:|:  :
Db         61 -------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIIN 113

Qy        113 DIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETST 172
              ||: |:   |:  |: ::   :|:          :   |  ::|| ||   : : ||
Db        114 GIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETID 173

Qy        173 FKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELN 232
                : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db        174 IGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELI 232

Qy        233 HAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSAR 286
              | :| ||||       :||::         |: |    ::  |:| |||    :|  |
Db        233 HVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIITPSTD 284

Qy        287 KYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNK 346
              |   :| |  :| |  ||| :     |  |  |  || |   ||:|| :| |: :::
Db        285 KSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVES 343

Qy        347 FVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSN 405
              | :|| |  ||| | |: | :: |  | |:   ||   |:||: :| |: ||||   :
Db        344 FDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKD 403

Qy        406 LNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLP 464
              :   : ||| : |  |  ::: |::     : | |::    :      | :  | | ||
Db        404 MEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDLF 454

Qy        465 FIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES- 523
              || | :    |:  : |   |:  |  ::  ::::|| :|   :::|  || ::||
Db        455 FIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELIL--DTDLISKIEL--PSENTESL 510

Qy        524 ----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALD 573
                       |  |   ::| |  |        : || ||    :: | : | | ::||
Db        511 TDFNVDVPVYEKQPAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDALL 563

Qy        574 NSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAII 629
              | |||::|      ||||   |: |||  |  :| ||     : :|:|||:|:| |:
Db        564 FSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIV 620

Qy        630 PYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKT 689
              |||| |||: |   :||| || : | :|||| ||  || :|||:: | :  :|:||||
Db        621 PYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIKT 680

Qy        690 IDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKY 749
              ||| | :| ::| | |  ::  ||| : |||  |   || :||||| |::  |   |  |
Db        681 IDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIY 740

Qy        750 SGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRN 809
              |  :| ||     :: : |:  |::|::||| ||  |||:|| | |:|  :::| :||
Db        741 SEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNT 800

Qy        810 TKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNIN 869
              |  |:| || : : |:|  :  |:|||  :  :||::  |||:::| :: |:| : |
Db        801 LKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEIL 860

Qy        870 DSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENI 929
              :: ||:|: : | |:| ||| |:|      |:||      ||| ||     |: ||||:||
Db        861 NNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA--NSKIRVTQNQNI 916

Qy        930 VYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLKQ 983
              ::||::  ||:|||||| |:    :  |     ||||: :|||||| | |  | :::||
Db        917 IFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLID 976

Qy        984 NEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINF 1042
                :|: | |:|  :   | |:|||||:||| : | ||||||||     :|::  :
Db        977 INGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIA 1035

Qy       1043 SKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKDY 1101
              :  | |:::          | |    :|:: | || ||   :|   :   |:  :||:
Db       1036 NGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDF 1086

Qy       1102 WGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN-------DFNEGYKIII 1150
              ||| | ||||||| |    | |:         :|:  :: | |          |  | ||
Db       1087 WGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFII 1146

Qy       1151 KRIRGNT---NDTRVRGGDILYFD-MTINNK----AYNLFMKNETMYADNHSTEDIYAIG 1202
              :| : |:   ||  ||  | :| |  :| :      | | | |          | ::
Db       1147 RR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE---------EKLFLAP 1196
```

```
Qy       1203 LREQTKDINDNIIFQIQPMNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRLGG----- 1256
             : :  : |     ||:  :    |: | :||   : |:   |  || :||   |
Db       1197 ISDSDEFYN---TIQIKEYDEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYESGIVFEE 1251

Qy       1257 --------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
                      ||     :  ||:  |   |      :| |:|  |
Db       1252 YKDYFCISKWY-----LKEVKRKPYNLKL---GCNWQFIPKDE 1286


RESULT 5
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
```

```
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190
```

Query Match 27.0%; Score 1823.5; DB 1; Length 1314;
Best Local Similarity 32.6%; Pred. No. 7.1e-72;
Matches 447; Conservative 253; Mismatches 526; Indels 147; Gaps 33;

```
Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLNK 60
              |||||||  |||||:|  |: ::      :   |||:||  ||::|:|:   : | : |
Db          1 PITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFNP 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
               |  :    | ||||||| |||||| ||: ::||| || :   || |: ::   ||: ||:
Db         61 PSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGNS 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKL---TN 177
               : :: || : :  | ::  :  :       ::  ::|| ||   :  |  |
Db        121 YSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDNK 180

Qy        178 NTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHN 237
               | |  ::|||::  ::  | :: |: |  ::       ||::  || |:||||| | :|
Db        181 NYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLHG 240

Qy        238 LYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYY 297
               |||: : : : | |       | |:   :   |:: |||   :||     :    || |: |
Db        241 LYGMQVSSHEIIPS--KQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDY 298

Qy        298 RSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQI 357
               ::|| :|: :|: |  : :  |  |||   :||:|  :|:|:  || :||  |||  :
Db        299 KAIANKLSQVTSCNDPNID--IDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYG 356

Qy        358 FTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLFMGQNLS 416
               |||    | :|:: |  | |  : ||   |:||| :|:   ||||   :|   : |||:
Db        357 FTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMR 416

Qy        417 RNP-ALRKVNPENMLYLFTKFCHKAID----GRSLYNKTLDCREL------LVKNTDLPF 465
               |  | | |:   ::       | | |       :|||:|   :|       :|| || |
Db        417 VNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTF 476

Qy        466 IGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEI 525
               | : :    : |  : ::  |:          | |:|::|:  |     |  :  |: |  : :
Db        477 IAEKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNL----QSKITLPN-DRTTPV 531

Qy        526 LPG-------ENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKV 578
               |        ::          |:|    | || :||    ::  | | |:::|| || |:
Db        532 TKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKI 591

Qy        579 YTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNI 638
               |:|||:: :||| | || ||| |  |:::|||     :| |:|||||| |:|||||||||
Db        592 YSYFPSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNI 651

Qy        639 SNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRI 698
                  |||  |   ||| :|||  || |:| : |   |      :: :|||||| ||:|
Db        652 VKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRY 711

Qy        699 KRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIK 758
               ::| : |: :   ||  : |||   ||||| || ||  |||  || ||| ||| ||| |  -
Db        712 EKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIA 771

Qy        759 SQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLI 818
               :: |||| |: | ::|| ||| |:|| | ::|    |: :   :| |||  :|  |:  |
Db        772 DEINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYI 831

Qy        819 DSHN-IILVGEVDKLKAKVNNSFQNTIPF----NIFSYTNNSLLKDIINEYFNNINDSKI 873
               :::  | : |: ||::|:|  |   |||     |:  : :|    |:|       :  | |
Db        832 KANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVI------LKKSTI 885

Qy        874 LSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGKVIVTQN 926
               |:|    : : | ||:|: |    | || |         | :|:       :  :||| :
Db        886 LNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSEVIVHKA 937

Qy        927 ENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWSIGIISN 975
               :| || |: :|::|||:|: |   |:|       |:|| |:|  :     ||||: :  |
Db        938 MDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGN 997

Qy        976 FLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGKLIDTIK 1033
               |::||| :    : | |  |: :     |  ||| |:|:||: : :  :|||| |: : :
Db        998 NLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAE 1056

Qy       1034 VKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQ 1093
               :  | |     || ::::         :::|  :  |  | || | |: |:|  |: |
Db       1057 ITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYL 1108
```

```
Qy       1094 YTNVVKDYWGNDLRYNKEYYMVNI-------------DYL---NRYMYANSRQIVFNTRR 1137
                 ::|:||| |||: |||:: :              ||:   |    | | :  ::  |
Db       1109 SITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRL 1168

Qy       1138 NNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLFMKNETM 1188
                |     | | ||||   |   |: |: || :   :: ||  :         | |   : :
Db       1169 YN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRI 1223

Qy       1189 YADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIG 1248
                ::   |       |  |             : :   |: |:    :: :| | :  :|
Db       1224 LRVGYNAPGIPLYKKMEAVK------------LRDLKTYSVQL--KLYDDKNAS-LGLVG 1268

Qy       1249 TYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
                |:  ::| |          ||  |:|   :              | |||  |
Db       1269 THNGQIGNDPNRDILIASNWY-FNHLKDKI-----------LGCDWYFVPTDE 1309


RESULT 6
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             26.8%;  Score 1810;  DB 1;  Length 1250;
  Best Local Similarity   33.0%;  Pred. No. 2.6e-71;
  Matches  442;  Conservative  261;  Mismatches  482;  Indels  154;  Gaps   42;

Qy          4 TINNFNYSDPVDNKNILYLDTHLNTLANEP------EKAFRITGNIWVIPDRFSRNSNPN 57
              |||:|||:|||:|: |||:          :|      |:| |  |||:||:|    : |
Db          2 TINSFNYNDPVNNRTILYI---------KPGGCQQFYKSFNIMKNIWIIPERNVIGTIPQ 52

Qy         58 LNKPPRVTSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDI 114
                 ||  || |:|   |||||||| :| :|| ||| : |:| |||    |  |:  ||
Db         53 DFLPP--TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKAN 110

Qy        115 PFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFK 174
              |: ||:|||    |  : | ::| ::    |:  :    : |:||| |   :: :  :|
Db        111 PYLGNDNTPDGDFIIN-DASAVPIQFSNGSQSI----LLPNVIIMGAEPDLFETNSSNIS 165

Qy        175 LTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHA 234
              | ||   :  |||:::|:: || :  : :          | :||  || | ||||| |:
Db        166 LRNNYMPSNHGFGSIAIVTFSPEYSFRFKDN---------SMNEFIQDPALTLMHELIHS 216

Qy        235 MHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKA 293
              :| |||   |    ||:   : :  :     :|  |   |||  :::|  :
Db        217 LHGLYGAKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNL 274

Qy        294 LDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNE 353
              |  |: || :|: :  :||       :  ||      ||    ::||  :|| ||| ::: :
Db        275 LADYKKIASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKK 329

Qy        354 LTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQ 413
              |   ||||: |  : |: |: |:        :|:|:|::|:|  |:||  :|| | | ||
Db        330 LYS-FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQ 386

Qy        414 NLSRNP-ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDV 472
              | : ||  :  :    ::    :||   :  : :  |:: | |  : | :| |:   :
Db        387 NANLNPRIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASENSY 442

Qy        473 KTDIFLRKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDSESE 524
              |      :||   |:    | |:          :|||||: |:           | : |
Db        443 NDD-----NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSDEKL 487
```

```
Qy        525 ILPGENQVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNS 575
Db        488 NLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQ 547

Qy        576 AKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGP 634
Db        548 PKIYTFFSSEFINNVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGL 607

Qy        635 ALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIKTID 691
Db        608 ALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAIN 667

Qy        692 NCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSG 751
Db        668 NALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTL 727

Qy        752 SDKENI--KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRN 809
Db        728 EEKNELTNKYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDEN 787

Qy        810 TKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNIN 869
Db        788 VKTYLLDYIIKHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIK 847

Qy        870 DSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENI 929
Db        848 SSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYI 905

Qy        930 VYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQN 984
Db        906 IYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDN 965

Qy        985 EDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFS 1043
Db        966 SGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVS 1025

Qy       1044 KTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWG 1103
Db       1026 DNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYNNEPNANILKDFWG 1076

Qy       1104 NDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RG 1155
Db       1077 NYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNS 1134

Qy       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215
Db       1135 STNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTNKEKTIKISSSGNRFNQVVV 1188

Qy       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQG 1271
Db       1189 M------NSVGNCTMNFKNN-NGNN------IGLLGFKADTVVASTWY------------ 1223

Qy       1272 NYASLLESTSTH---WGFV 1287
Db       1224 -YTHMRDNTNSNGFFWNFI 1241


RESULT 7
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             26.5%;  Score 1794;  DB 1;  Length 1250;
  Best Local Similarity   33.0%;  Pred. No. 1.3e-70;
  Matches  439;  Conservative  257;  Mismatches  495;  Indels  140;  Gaps   41;

Qy          5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
              ||:|||:|||::: |||:       |  |:| |  |||:||:|    : |    ||
Db          3 INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP-- 57

Qy         65 TSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNN 121
              || |:|   |||||||| :| :|| ||| : |:| |||:    |  |:  ||   |: ||:|
Db         58 TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDN 117

Qy        122 TPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFA 181
              || | |    | ::|::|   |:  :    : |:||| |   :: :  :|   | ||
Db        118 TPDNQFHIG-DASAVEIKFSNGSQDI----LLPNVIIMGAEPDLFETNSSNISLRNNYMP 172

Qy        182 AQEGFGALSIISISPRFMLTYS-NATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              :   ||:::|:: || :   :: | |           ||  || | ||||| |::| |||
Db        173 SNHRFGSIAIVTFSPEYSFRFNDNCMN----------EFIQDPALTLMHELIHSLHGLYG 222

Qy        241 I-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRS 299
              |    ||:   : :  :    :|  |   |||  :::|  :         | |:
Db        223 AKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNLLADYKK 280

Qy        300 IAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFT 359
              || :|: :  :||      :  ||      ||    ::||  :|| ||| ::: :|   ||
Db        281 IASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS-FT 334

Qy        360 EFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNP 419
              ||:    : |: |: |:        :|:|::|:|  |:||  :|| | | ||| : ||
Db        335 EFDLRTKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLNP 392

Qy        420 -ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFL 478
               :  :   ::    :||  :  : :  |:: | |  : :| :|  :    |
Db        393 RIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASENSYNDD--- 445

Qy        479 RKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDSESEILPGEN 530
              :||   |:    | |:         :|||||: |:          | :  |   | :|
Db        446 --NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSDEKLNLTIQN 493

Qy        531 QVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTY 581
                :   ||:         :|: || ::||::||: :   :   | ||: ||    |:||:
Db        494 DAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTF 553

Qy        582 FPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISN 640
              | :   | ||  ||  ||: |   |: ||||   :| |:|||:|:| ::|||| |||| |
Db        554 FSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGN 613

Qy        641 SVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIKTIDNCLEQR 697
              ::||| :|  : |  |||| ||  || :  | | |       :|::|| |:| |::|
Db        614 EAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKER 673

Qy        698 IKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENI 757
              ::||: | :::  |:::| ||||    ||| :|  | |||  |: :|  |:  :|  :
Db        674 DEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNEL 733

Qy        758 --KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLI 815
                |  :: ::| |: |:| |||||::|:| |::|| | :    |::| |:| | |  |:
Db        734 TNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLL 793

Qy        816 NLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILS 875
              |  |  ||     :| : |  ::  |:||| : |||:: :|    |::|  |  | :|:
Db        794 NYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLN 853

Qy        876 LQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMY 935
              :: : :  ||||||:: ::  |||    |      |  |   :  :| ::||: |:|:: |
Db        854 MRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYIIYDNKY 911

Qy        936 ESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDSEQS 990
              ::||||||:||    || |:    ||||: :: |||||  : :  | :::| : |    |
Db        912 KNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTFEDNRGINQK 971

Qy        991 INFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFE 1049
              : |:|  :|     | ||| |||:||: :|: |:|||| |||   :  |  |: |  | |:
Db        972 LAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFK 1031

Qy       1050 INKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYN 1109
              |      |  |              || | || |||  :|  |:::   ||::||:||| | |:
Db       1032 IVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYD 1082

Qy       1110 KEYYMVNIDYLNRYM--YANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RGNTNDTRVR 1163
              ||||::|:   | ::    :|   : | |       |    | |: |:|:   :||| ||
Db       1083 KEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVR 1142

Qy       1164 GGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNN 1223
                | :|  :   :  :||     :|||  :|     | :       |  ::       |
Db       1143 KNDQVYINFVASKT--HLF----PLYADTATTNKEKTIKISSSGNRFNQVVVM------N 1190

Qy       1224 TYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYASLLES 1279
              :    :  ||:| || |      ||   |:  :  ||           |  : :
Db       1191 SVGNCTMNFKNN-NGNN------IGLLGFKADTVVASTWY-------------YTHMRDH 1230

Qy       1280 TSTH---WGFV 1287
              |:::   | |:
Db       1231 TNSNGCFWNFI 1241


RESULT 8
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
```

```
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match           26.2%;  Score 1772;  DB 1;  Length 1295;
  Best Local Similarity  32.8%;  Pred. No. 1.2e-69;
  Matches  443;  Conservative  243;  Mismatches  542;  Indels  122;  Gaps   34;

Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNLN 59
              |      ||| |||:  :| |:    |    :| |||:|   |||||:|  |:     :||
Db          1 PFVNKQFNYKDPVNGVDIAYIKIP-NVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN 59

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
               ||          |||  |||||::|| :|| : |||:|| | ::|  |:  :   ||| |
Db         60 PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG 119

Qy        120 NNTPINTFDFD---VDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF--- 173
                   :| | :   :| | ::|    |:   ::  :|  ::| ||  :||  |  :|
Db        120 -----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQFECKSFGHE 170

Qy        174 --KLTNNTFAAQEGFGALSIISISPRFMLTYSNA----TND-VGEGRFSKSEFCMDPILI 226
                ||  |       |:|:   |  || |    :  :     ||  :| |:|:       || :
Db        171 VLNLTRN------GYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA-----TDPAVT 219

Qy        227 LMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSAR 286
              | ||| || | |||||| |   :  | :| :|    ::: : |:  |||     |
Db        220 LAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQE 278

Qy        287 KYFEEKALDYYRSIAKRLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVVESSGEV 340
                |      : :: ||  ||        |||:         :| : : : ||    ::||:
Db        279 NEFRLYYYNKFKDIASTLNKAKSIVGTTASL--------QYMKNVFKEKYLLSEDTSGKF 330

Qy        341 TVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFN 400
              :|:: || :||  ||:|:||| |: | : | ||| ||:         ||:    | | :|||
Db        331 SVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFN 390

Qy        401 IPKSNLNVLFMGQNLS-RNPALRKVNPENMLYLFTK-FCHKAI---DGRSL---YNKTLD 452
              :  :||   | |||     |    |:      |: | |  | : |     :||   ||| |:
Db        391 LRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALN 450

Qy        453 CRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS---KNT 506
                : | | || |     :   |:   ::|  :| :    :|:|:|   |  |:    |
Db        451 DLCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDNE 510

Qy        507 SEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTR 566
               |:  :: |   |  : |::|   : | : :   :|    ::|| :|:           |
Db        511 PENISIENLSSDIIGQLELMPNIER-FPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTN 569

Qy        567 SIEEALDNSAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDV 625
              |: ||| | ::|||:| :    |||   :  :|| |   :| |||        | |||:|:
Db        570 SVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADI 629

Qy        626 SAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNE 685
              : |||||||||| | : : :|  |   :|  ||||  ||  || || | : | :  :
Db        630 TIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVL 689

Qy        686 IIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLE 745
               ::|||| | :| ::| : |:::  ||::: || : |   :| ::|  || | || |: :
Db        690 TVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ 749

Qy        746 YKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNE 805
              | :|:  :| ||   :::| : |:  |::|| |||||: :|||:||  :|:|  :  | :
Db        750 YNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLED 809

Qy        806 FDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYF 865
              || : |  |:  |  :   |:|:||:|| ||||:    ||| :  | :|  |     ||
Db        810 FDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYI 869

Qy        866 NNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVI 922
              || :: ||:|:   | |:| | | :::    |  :||       | | ||     |:
Db        870 KNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS-----KIE 924

Qy        923 VTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVF 979
              |      |||||||:|| ||||||  |: :::      ||||: ::|||||  : :    :::
Db        925 VILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIW 984

Qy        980 TLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELT 1038
              ||:  :: :| : | |    |   | |:| |||:||| : | ||||||:|||   :  |
Db        985 TLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG 1044

Qy       1039 GINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVV 1098
              |: |  | |:::    ||         :  :||: | :| |||: |:|  |:::   : ::
Db       1045 NIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1096

Qy       1099 KDYWGNDLRYNKEYYMVNIDYLNRYMYANS-----------RQIVFNTR-RNNNDFNEG 1145
              ||:||: |:|:|  |||:|:    |:|:  |:             |  |  |    |:   |
Db       1097 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy       1146 YKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLRE 1205
              | |||:       |  ||  | :| :: : || | |         | | |   :  |
Db       1157 TKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVEKILSAL 1208

Qy       1206 QTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLV 1265
              :  |: :      :     ::  | |    | : |  || ::|          | :
Db       1209 EIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---------NNIA 1258

Qy       1266 PTVKQGNYASLLESTS----THWGFVPVSE 1291
                |    :| :|     | |:|| :
Db       1259 KLVASNWYNRQIERSSRTLGCSWEFIPVDD 1288


RESULT 9
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
```

```
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             26.2%;  Score 1768;  DB 1;  Length 1274;
  Best Local Similarity   33.3%;  Pred. No. 1.7e-69;
  Matches  446;  Conservative  251;  Mismatches  519;  Indels  124;  Gaps   41;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||: ||:|||:|||::  |||:       : :  ||| |  |:|:||:| :  :|| : :
Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTNPSDFD 60

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
                |  : :  | |||||||:||::|| :||  |||||||||   |: |:  :|    |: ||
Db         61 PPASLKNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYLGN 120

Qy        120 NNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPR---------ENIIDPET 170
              ::|||: |       ||::|     |   :  :|  |:  ||             :|||:
Db        121 DHTPIDEFSPVTRTTSVNIKL--STNVESSMLLNLLVLGAGPDIFESCCYPVRKLIDPDV 178

Qy        171 STFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSE-FCMDPILILMH 229
              :  :|       |||:::|:: || :  |:     ||:  |  | :|  |  || : | |
Db        179 -VYDPSN------YGFGSINIVTFSPEYEYTF----NDISGGHNSSTESFIADPAISLAH 227

Qy        230 ELNHAMHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKY 288
              || ||:| |||   :  ::||    : :  ::  ::|| |   |||  :::|  : ::
Db        228 ELIHALHGLYGARGVTYEETIEVKQAPLMIAEKPIRLE--EFLTFGGQDLNIITSAMKEK 285

Qy        289 FEEKALDYYRSIAKRLNSITTANPSSFNKY-IGEYKQKLIRKYRFVVESSGEVTVNRNKF 347
                |  | || ||: : :| |     :| | |||      ||      :  | ||| |||
Db        286 IYNNLLANYEKIATRLSEVNSAPP----EYDINEYKDYFQWKYGLDKNADGSYTVNENKF 341

Qy        348 VELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLN 407
              |:| :|   ||| :  |  : |: |  |           |:|||::| :  ||||   ||
Db        342 NEIYKKLYS-FTESDLANKFKVKCRNTYFIKYEFLKVPNLLDDDIYTVSEGFNI--GNLA 398

Qy        408 VLFMGQNLSRNPALRKVNPE-NMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFI 466
              |  ||::  || :   |:  ::    |||   |  :          |  : | |::| |:
Db        399 VNNRGQSIKLNPKIIDSIPDKGLVEKIVKFCKSVIPRKGTKAPPRLC--IRVNNSELFFV 456

Qy        467 GDISDV-KTDIFLRKDINEETEV-IYYPDNVSVDQVILSKNTSEHGQL----------DL 514
              |   : ||   |:|:: | :   |  :|  :|:|||  |:    |:           |
Db        457 ASESSYNENDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSQTIPQISNRTLNTLVQDN 514

Qy        515 LY-PSIDSESEILPGENQV-FYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
              | |  ||       | :::  ||      ||: | ::|| :||: :    : : | ||: ||
Db        515 SYVPRYDSN-----GTSEIEEYD----VVDF-NVFFYLHAQKVPEGETNISLTSSIDTAL 564

Qy        573 DNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYI 632
              :|   :    : :|  |  ||: | : |: ||||   :| |:|||:|:|  |:||:
Db        565 LEESKDIFFSSEFIDTINKPVNAALFIDWISKVIRDFTTEATQKSTVDKIADISLIVPYV 624

Qy        633 GPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKV---QERNEIIKT 689
              | ||||     :||| ||| : || ||||  || ||| :  | | | :   : :|: ||
```

```
Db        625 GLALNIIIEAEKGNFEEAFELLGVGILLEFVPELTIPVILVFTIKSYIDSYENKNKAIKA 684

Qy        690 IDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKY 749

Db        685 INNSLIEREAKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTAIEYKYNNY 744

Qy        750 SGSDKENIKSQ--VENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFD 807

Db        745 TSDEKNRLESEYNINNIEEELNKKVSLAMKNIERFMTESSISYLMKLINEAKVGKLKKYD 804

Qy        808 RNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNN 867

Db        805 NHVKSDLLNYILDHRSILGEQTNELSDLVTSTLNSSIPFELSSYTNDKILIIYFNRLYKK 864

Qy        868 INDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNE 927

Db        865 IKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQF--GIYNSRLSEVNIAQNN 922

Qy        928 NIVYNSMYESFSISFWIRINKW---VSNLPGYTIIDSV-KNNSGWSIG---IISNFLVFT 980

Db        923 DIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWT 982

Qy        981 LKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTG 1039

Db        983 LQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGD 1042

Qy       1040 INFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVK 1099

Db       1043 IHVSDNILFK---------IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILK 1093

Qy       1100 DYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN--------NNDFNEGYKIIIK 1151

Db       1094 NYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVIIR 1153

Qy       1152 R---IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTK 1208

Db       1154 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL-------YADTKSEKEKII-----RTS 1201

Qy       1209 DINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYL 1264

Db       1202 NLNDS-LGQIIVMDSIGNNCTMNFQNN-NGSN------IGLLGFHSNNLVASSWYYNNIR 1253

Qy       1265 VPTVKQGNYASLLESTSTHW 1284

Db       1254 RNTSSNGCFWSSI-SKENGW 1272
```

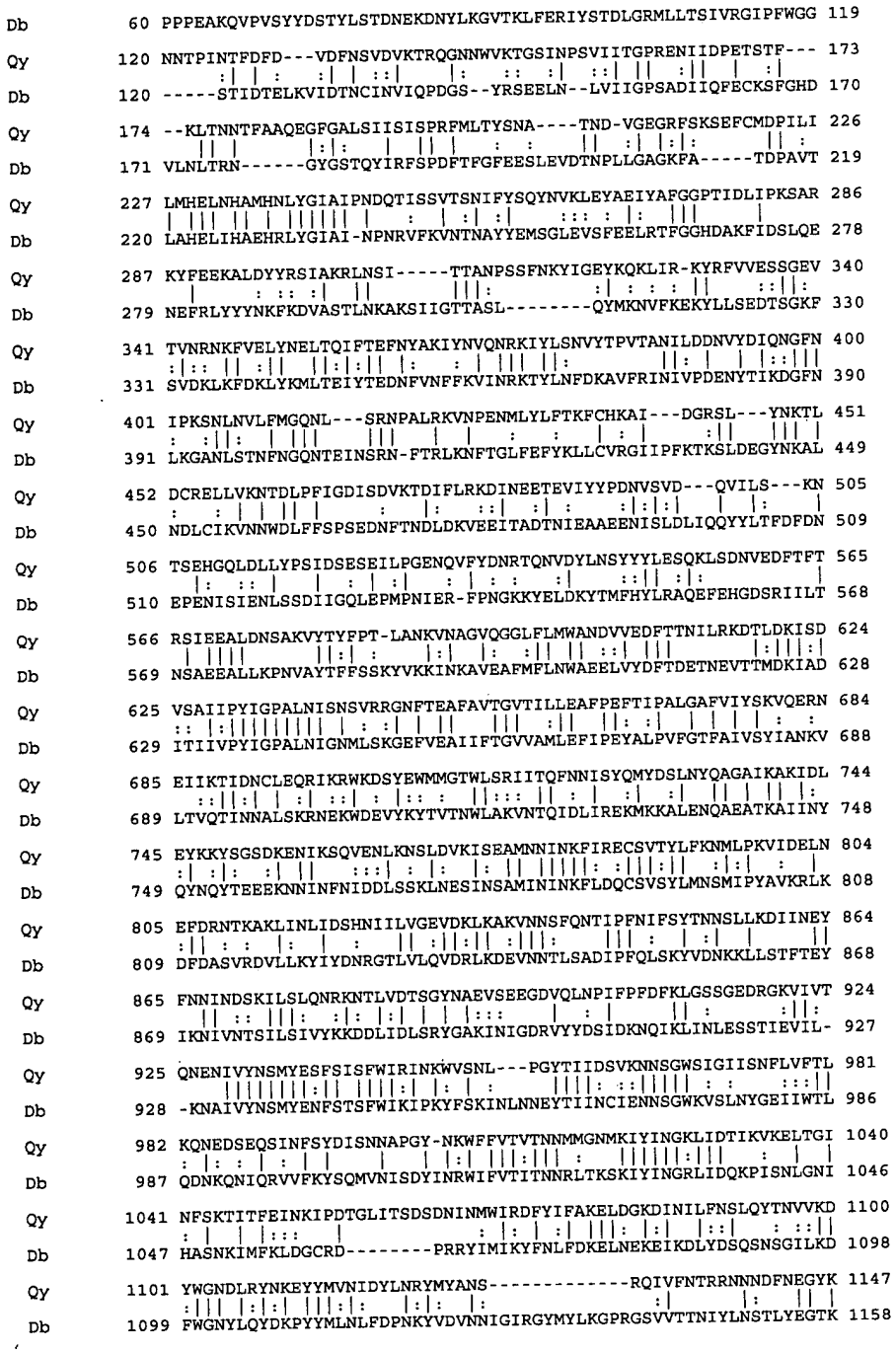
RESULT 10
BXA2_CLOBO
ID BXA2_CLOBO STANDARD; PRT; 1295 AA.
AC Q45894; P77780;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kyoto-F;
RX MEDLINE=94143603; PubMed=8310180;
RA Willems A., East A.K., Lawson P.A., Collins M.D.;
RT "Sequence of the gene coding for the neurotoxin of Clostridium
RT botulinum type A associated with infant botulism: comparison with
RT other clostridial neurotoxins.";
RL Res. Microbiol. 144:547-556(1993).
RN [2]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=Kyoto-F;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC   binds with high affinity to peripheral neuronal presynaptic
CC   membrane, is then internalized by receptor-mediated endocytosis.
CC   The C-terminus of the heavy chain (H) is responsible for the
CC   adherence of the toxin to the cell surface while the N-terminus
CC   mediates transport of the light chain from the endocytic vesicle
CC   to the cytosol. After translocation, the light chain (L)
CC   hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC   neurotransmitter release. Inhibition of acetylcholine release
CC   results in flaccid paralysis, with frequent heart or respiratory
CC   failure (By similarity).
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC   neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC   detected action on small molecule substrates.
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC   heavy chain (H) (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC   botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X73423; CAA51824.1; -.
DR EMBL; X87974; CAA61234.1; -.
DR PIR; I40645; I40645.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET   0   0   BY SIMILARITY.
FT CHAIN   1   447   BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN   448   1295   BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL   222   222   ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE   223   223   BY SIMILARITY.
FT METAL   226   226   ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID   429   453   INTERCHAIN (BY SIMILARITY).
FT DISULFID   1234   1279   BY SIMILARITY.
FT TRANSMEM   626   646   POTENTIAL.
FT TRANSMEM   655   675   POTENTIAL.
SQ SEQUENCE   1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

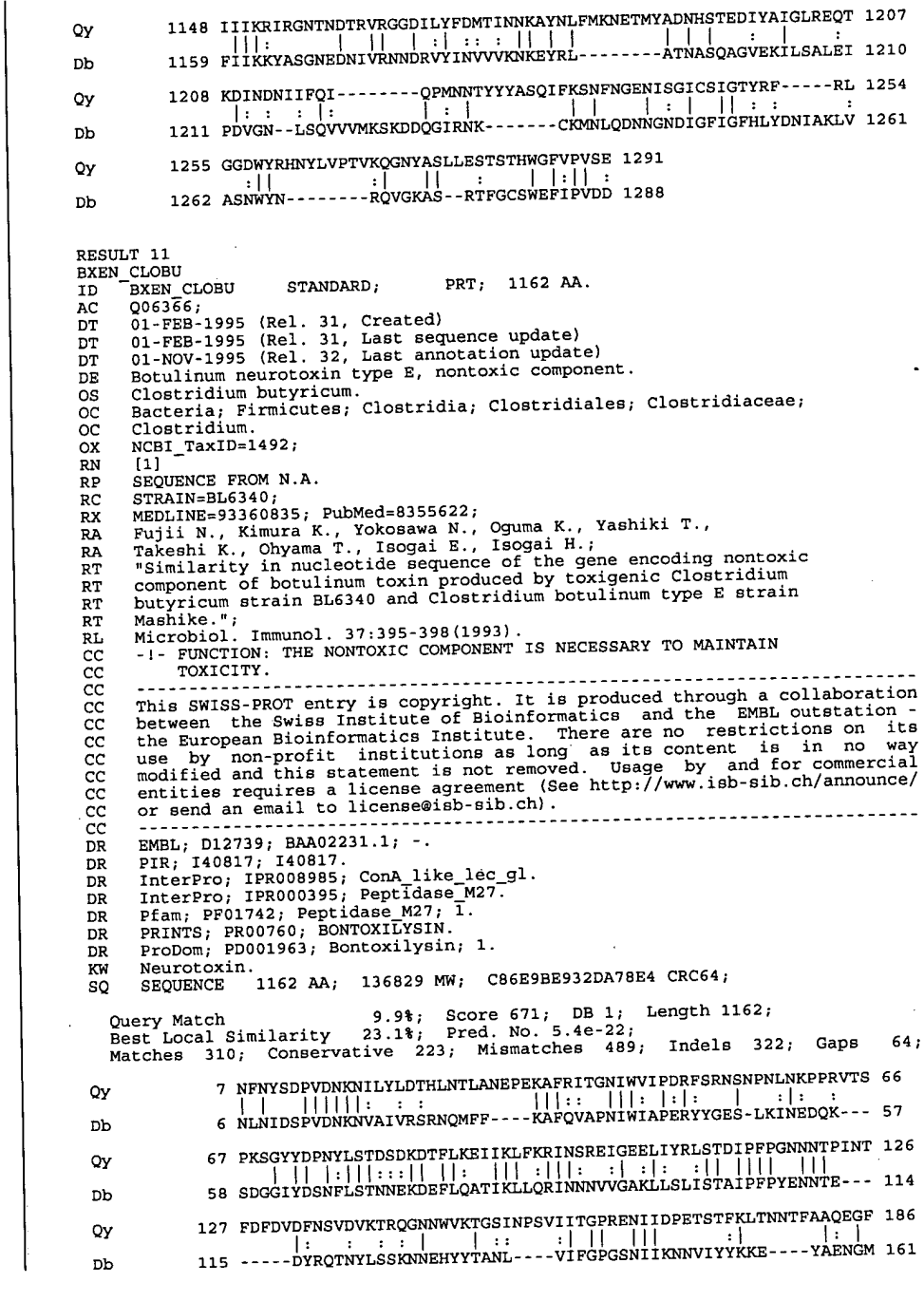
  Query Match   25.4%; Score 1717; DB 1; Length 1295;
  Best Local Similarity   32.2%; Pred. No. 2.9e-67;
  Matches 437; Conservative 244; Mismatches 540; Indels 136; Gaps   35;

```
Qy          2 PITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDR--FSRNSNPNLN 59

Db          1 PFVNKQFNYKDPVNGVDIAYIKIP-NAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN 59

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
```





RESULT 11
BXEN_CLOBU
ID BXEN_CLOBU STANDARD; PRT; 1162 AA.
AC Q06366;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=93360835; PubMed=8355622;
RA Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT "Similarity in nucleotide sequence of the gene encoding nontoxic
RT component of botulinum toxin produced by toxigenic Clostridium
RT butyricum strain BL6340 and Clostridium botulinum type E strain
RT Mashike.";
RL Microbiol. Immunol. 37:395-398(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D12739; BAA02231.1; -.
DR PIR; I40817; I40817.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136829 MW; C86E9BE932DA78E4 CRC64;

Query Match 9.9%; Score 671; DB 1; Length 1162;
Best Local Similarity 23.1%; Pred. No. 5.4e-22;
Matches 310; Conservative 223; Mismatches 489; Indels 322; Gaps 64;



```
Qy       187 GALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPND 246
Db       162 GTMLEIWFQP--FLTH-----------KYDEFYVDPALELIKCLIKSLYYLYGIK-PND 206

Qy       247 QTISSVTSNIFY---SQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKR 303
Db       207 NL------NIPYRLRNEFN-SLEYSE--------LDMI-----DFLISGGIDY-----KL 241

Qy       304 LNSITTANPSSF-NKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFN 362
Db       242 LNT----NPYWFIDKYFID-TSKNFEKYK----NDYEIKIKNNNYIA-----------N 280

Qy       363 YAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALR 422
Db       281 SIKLYLEQKFKINVKDIWE------LNLSYFSKEFQIMMPERYNNAL--------NHYYR 326

Qy       423 KVNPENMLYLFTKFCHKAIDG-----------RSLYNK-TLDCRELLV-----KNTDL-- 463
Db       327 K-----EFYVIDYFKNYNINGFKNGQIKTKLPLSKYNKEIINKPELIVNLINQNNTVLMK 381

Qy       464 -PFIGD-----ISDVKTDIFLRKDINEETEVIY-YPDNVSVDQV--ILSKNTSEHGQLDL 514
Db       382 SNFYGDGLKGNVDNFYSNYIIPYNLNYEHSINYSYLDNVNIEEIEKIPPINDED------ 435

Qy       515 LYPSIDSESEILPGENQVFYDNRTQNVDYLN--SYYYLESQKLSDN----VEDFTFTRSI 568
Db       436 IYPYRKNADTFIP----VYNITKAKEINTTTPLPVNYLQAQMIDSNDINLSSDFLKVISS 491

Qy       569 EEALD----NSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD 624
Db       492 KGSLVYSFLNNTMDYLEFIKYDKPIDTDKK---YYKWLKAIFRNYSLDITETQEISNQFG 548

Qy       625 VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPAL------------- 671
Db       549 DTKIIPWIGRALNILNT--NNSFVEEFKNLGPIFLINKKENITIPKIKIDEIPSSMLNFS 606

Qy       672 ------GAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISY 725
Db       607 FKDLSENLFNIYCK---NNFYLKKI-------------YYNFLDQWWTQYYSQYFDLIC 649

Qy       726 QMYDSLNYQAGAIKAKID------LEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNI 779
Db       650 MASKSVLAQEKLIKKLIQKQLRYLMENSNISSTNLILINLTTTNTLRDISNQSQIAINNI 709

Qy       780 NKFIRECSVTYLFKNMLPKVIDELNE----FDRNTKAKLINLIDSHNIILVGEVDKLKAK 835
Db       710 DKFFNNAAMCVFENNIYPKFTSFMEQCIKNINKSTKEFILKCTN------INETEKSHLI 763

Qy       836 VNNSFQNTIPF---------NIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLV-D 885
Db       764 MQNSFSN-LDFDFLDIQNMKNLFNLYTELLIKEQTSPY-----ELSLYAFQEQDNNVIGD 817

Qy       886 TSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVY------NSMYESFS 939
Db       818 TSGKNTLVE----------YPKDIGLVYGINNNAIHLTGANQNIKFTNDYFENGLTNNFS 867

Qy       940 ISFWIRINKWVSNLPGYTI----IDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSY 995
Db       868 IYFWLR------NLNQNTIKSKLIGSKEDNCGWEIYFENNGLVFNIIDSNGNEKNIYLS- 920

Qy       996 DISNNAPGYNKWFFVTVT-NNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIP 1054
Db       921 NISNKS-----WHYIVISINRLKDQLLIFIDNILVANEDIKEILNIYSS---------- 964

Qy      1055 DTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYM 1114
Db       965 --DIISLLSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNRTYEL 1022

Qy      1115 VNIDY----LNRYMYANSRQIVFNTRRNNNDFN---EGYKIIIKRIRGNTNDTR------ 1161
Db       1023 FNYVFPEIAINKIEQNNN---IYLSNNNENSLNFKPLKFKLL------NTNPNKQYVQKW 1073

Qy       1162 ------VRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215
Db       1074 DEVIFSVLDGTEKYLDISIDN--------NRIQLVDNKNNAKTFIIN--------NDIFI 1117

Qy       1216 FQIQPM--NNTYYYASQIFKSNFN 1237
Db       1118 SNCLTLTYNNVNVYLS-IKNQDYN 1140
```

RESULT 12
BXEN_CLOBO
ID BXEN_CLOBO STANDARD; PRT; 1162 AA.
AC P46082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
GN ENT-120.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mashike;
RX MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
RT component of Clostridium botulinum type E progenitor toxin.";
RL J. Gen. Microbiol. 139:79-86(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D12697; BAA02194.1; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDDAE0F39D CRC64;

Query Match 9.8%; Score 665; DB 1; Length 1162;
Best Local Similarity 22.9%; Pred. No. 9.8e-22;
Matches 307; Conservative 221; Mismatches 497; Indels 316; Gaps 61;

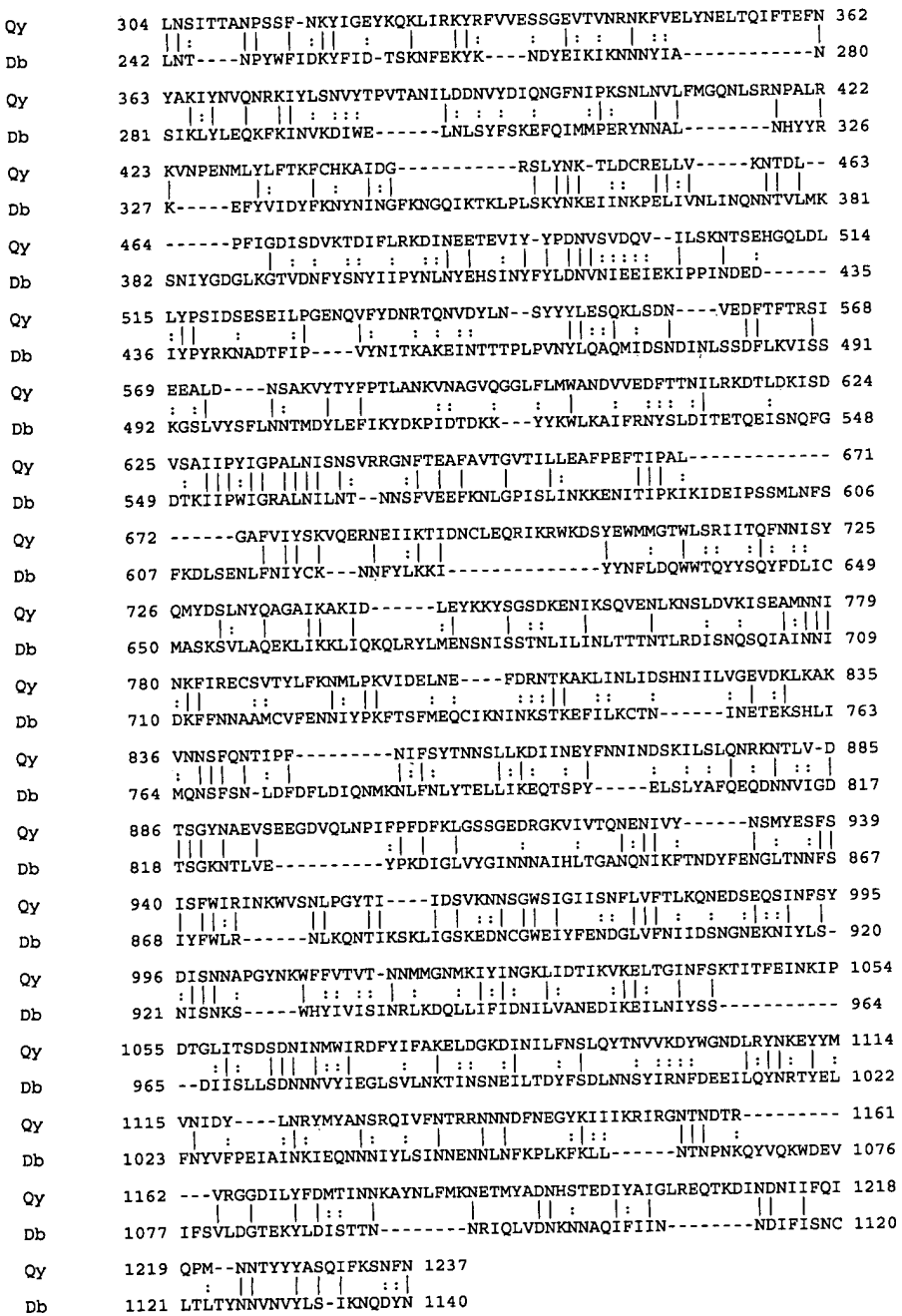
```
Qy         7 NFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTS 66
Db         6 NLNIDSPVDNKNVAIVRSRNQMFF----KAFQVAPNIWIVPERYYGES-LKINEDQKF-- 58

Qy        67 PKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINT 126
Db        59 -DGGIYDSNFLSTNNEKDDFLQATIKLLQRINNNVVGAKLLSLISTAIPFPYENNTE--- 114

Qy       127 FDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGF 186
Db       115 -----DYRQTNYLSSKNNEHYYTANL----VIFGPGSNIIKNNVIYYKKE----YAESGM 161

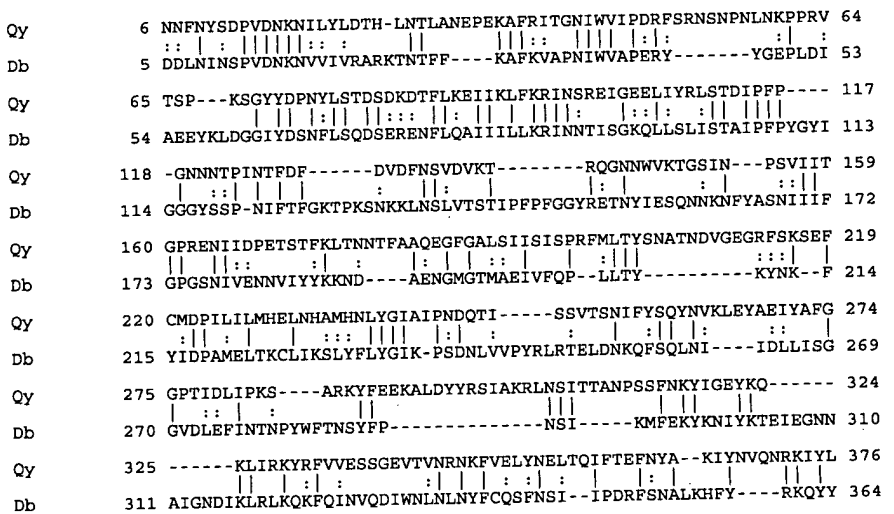
Qy       187 GALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPND 246
Db       162 GTMLEIWFQP--FLTH-----------KYDEFYVDPALELIKCLIKSLYYLYGIK-PND 206

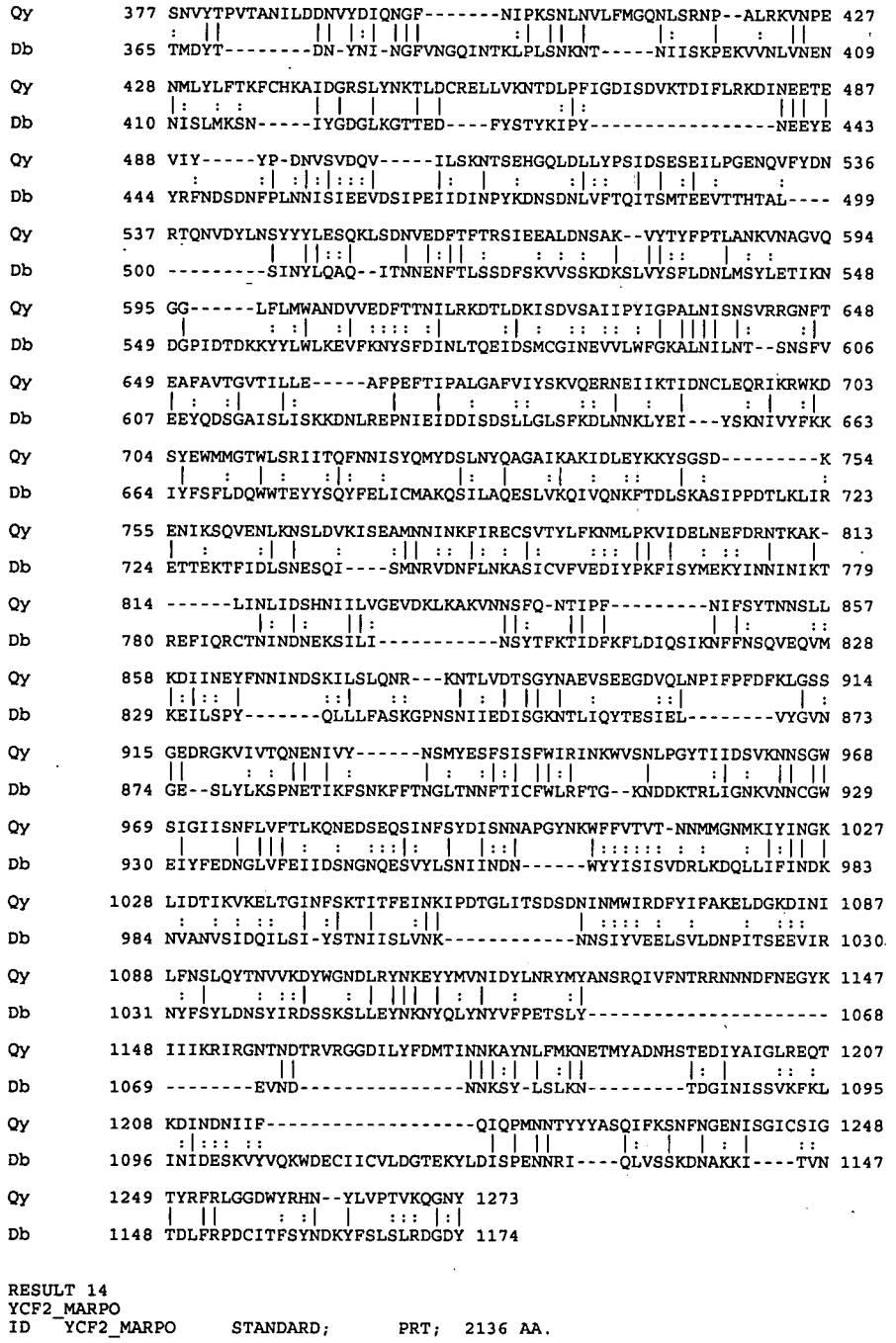
Qy       247 QTISSVTSNIFY---SQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKR 303
Db       207 NL------NIPYRLRNEFN-SLEYSEL-----NMIDFLISG--------GIDY-----KL 241
```



```
RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             9.5%;  Score 639;  DB 1;  Length 1196;
  Best Local Similarity   21.3%;  Pred. No. 1.4e-20;
  Matches  299;  Conservative  247;  Mismatches  485;  Indels  376;  Gaps   66;
```





RESULT 14
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT; 2136 AA.

AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC -!- SIMILARITY: Belongs to the ycf2 family.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; X04465; CAA28078.1; -.
DR PIR; S01591; A05037.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_centr.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 4.1%; Score 275; DB 1; Length 2136;
Best Local Similarity 19.6%; Pred. No. 0.00017;
Matches 285; Conservative 195; Mismatches 481; Indels 494; Gaps 74;



```
Db        775 LNFQEIQKILNCFSLFFNSKNIKKTKIFKNSYFINENLTTTFSFNDKEFNIFFLELFISE 834

Qy        424 VNPENMLYLFTKFCHKAI--DGRSLYNKTLDCRELL----VKNTDLPFIGDISDVKTDIF 477

Db        835 INNDFLMRFFKKYLYYRIYKDKEILFN-PIENRQLLQNFFEKTKILTFIDFLQD------ 887

Qy        478 LRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSI------DSESEILPGE-- 529

Db        888 --PELNYNNRFIFH----------LEKKTIKNN--NLLYLRLLKIFLKDKRNFLLINEIK 933

Qy        530 ------NQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYF- 582

Db        934 SFIEKKNNLFIKSQLSNVLLVKNSY-----KFFDNIFNFHFLKQKEKNIEIILNNQNYFE 988

Qy        583 -----PTLANKVNAGVQGGLFLMWANDVVEDFT--------TNILRKDTLDKISDVSAII 629

Db        989 KSLLKKTYLKNLN-----------LNNSYSKFSYKIFIFQLLNILNKNNYKTFQWISELI 1037

Qy        630 PYIGPALN--ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSK-VQERNEI 686

Db       1038 FY-SKNLNYKIQNKIEKNNY--------------------------CYNKNISYKKKK 1068

Qy        687 IKTID----NCLEQRIKRWKDSYEWM-MGTWLSRIITQFNNISYQMYDSLNYQAGAIKAK 741

Db       1069 IKTVNFFEKNNLFQTNNSWFFTLEWWEYNTYILLQIIQ--ETFFQITDVLEY----FKKK 1122

Qy        742 --IDLEYKKYSGSDKENIKS-QVENLK--------NSLDVKISEAMNNI---NKFIRECS 787

Db       1123 KIIEKNLKFFLKSKKISLKTLSFHNFKLKWNLRFFNEINYKKNYLLNFLWSDFNLINNCN 1182

Qy        788 VTY---------------------------LFKNMLPKVIDELNEFDRNTKAKLINLID 819

Db       1183 NLYWVIFSLVIFIFLYYQKIFSIIIGSDCFHLWKNF--EIIQYLT--DRSRSLYFTKLTR 1238

Qy        820 SHNIILVGEVDKLKAKVNNSFQNTIPF--NIFSY--TNNSLLKDIINEYFNNINDSKILS 875

Db       1239 RNKTAL----NKTENLLSYFFQNLTHYITNIKFYLLTKKNLKKWLIN--------NKTLD 1286

Qy        876 LQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVI---VTQNENIVY- 931

Db       1287 LSRRKRKLL----------VQSLITHNKIQNYGFELNSNKQFFTSYFGYQITNQQGLLYF 1336

Qy        932 ---------NSMYESFSISFWIRINKWVSNLPGYTIIDSVK----NN------------- 965

Db       1337 QYLAQFFQKNLINNSLDLA-----NKWIVFSFWHKIFSSQKLRQTNNIELGFQNIPVPLQ 1391

Qy        966 --SGWSIGII--------SNFLVFTLKQNEDSEQSINFSYDISNNAPGYNK-------WF 1008

Db       1392 FGLSYSKGILLIGPIETGRSYLIKNL-----AAESYVPLFKISINKLLYNKPDVITESWM 1446

Qy       1009 FVTVTNNMMGNMKIYINGKL---------IDTIKVKELTGINFSKTITF----------- 1048

Db       1447 NILIESLRRLNLTLDFAKKMSPCIIWIQNIHQLNVNRLTQ-NVESDPTFLLGILLKYFQT 1505

Qy       1049 EINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINI-LFNSLQYTNVVKDYWGNDLR 1107

Db       1506 DFSKTKKNNIIVIGSTHLPKKVDPALISPNRLD-KIINVRLFNISQRK---KQF---PLL 1558

Qy       1108 YNKEYYMV--NIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGG 1165

Db       1559 LKKKNFQLKENLFFLNEF---GSRTMGYNLRDLSALTNEVLLISITKNRSFIDTDTLK-- 1613

Qy       1166 DILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMN--- 1222

Db       1614 ------LAFHRQIFGLTYTNNKLNFDRIFKIVIYKVG-----KTIIQNILIKSSSMNLLN 1662

Qy       1223 -------NTYYYASQ 1230

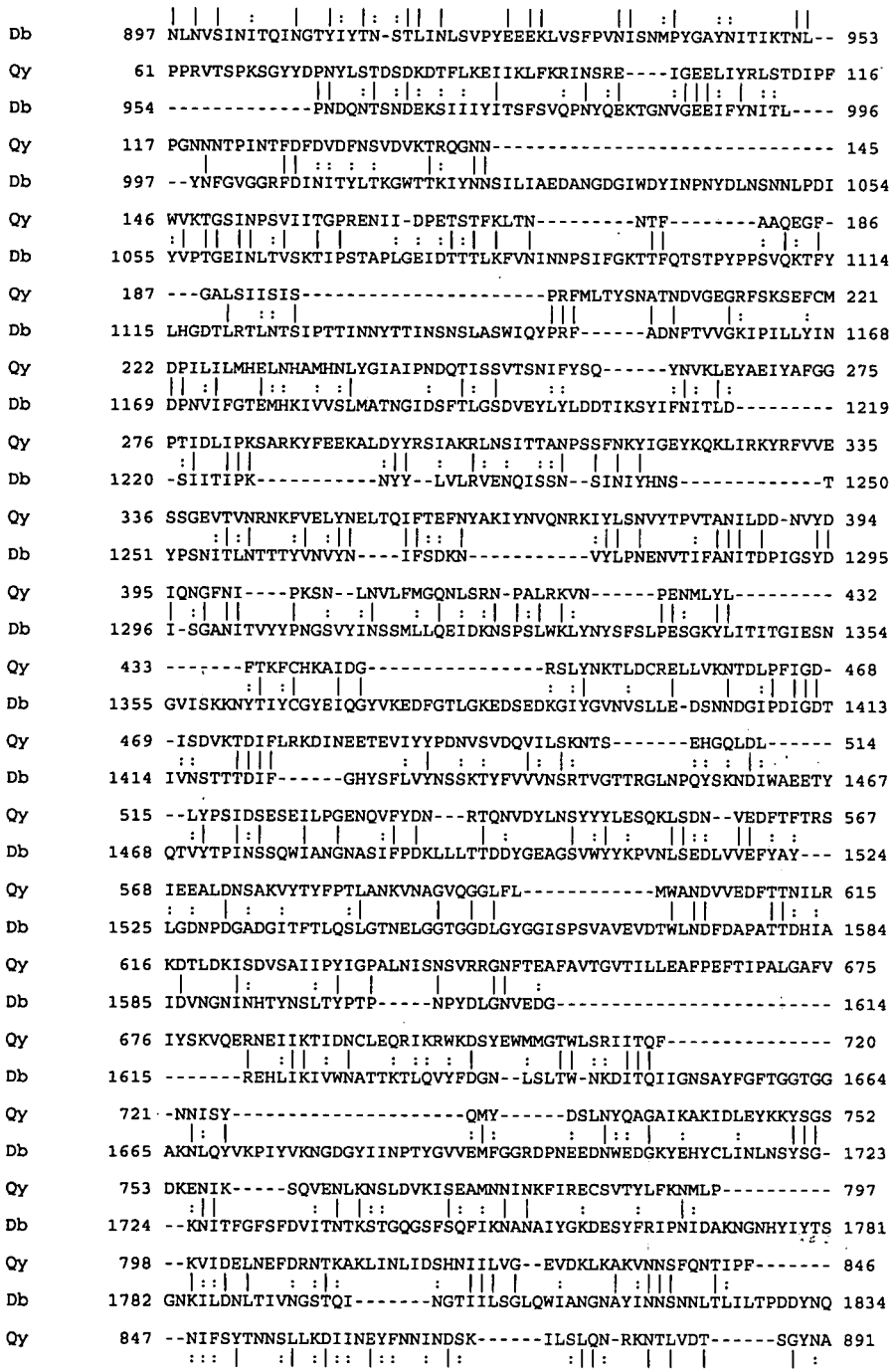
Db       1663 IGNFLWKKNFYYLSK 1677

RESULT 15
YD96_METJA
```

```
ID   YD96_METJA     STANDARD;      PRT;  2894 AA.
AC   Q58791;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hypothetical protein MJ1396.
GN   MJ1396.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: Contains 20 PbH1 repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67579; AAB99406.1; -.
DR   PIR; C64474; C64474.
DR   TIGR; MJ1396; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001220; Lectin_legB.
DR   InterPro; IPR006626; PbH1.
DR   Pfam; PF00139; lectin_legB; 1.
DR   SMART; SM00710; PbH1; 20.
KW   Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   REPEAT      543    567       PBH1 1.
FT   REPEAT     2085   2107       PBH1 2.
FT   REPEAT     2135   2156       PBH1 3.
FT   REPEAT     2158   2180       PBH1 4.
FT   REPEAT     2201   2223       PBH1 5.
FT   REPEAT     2224   2244       PBH1 6.
FT   REPEAT     2245   2266       PBH1 7.
FT   REPEAT     2267   2289       PBH1 8.
FT   REPEAT     2290   2311       PBH1 9.
FT   REPEAT     2341   2363       PBH1 10.
FT   REPEAT     2367   2389       PBH1 11.
FT   REPEAT     2390   2419       PBH1 12.
FT   REPEAT     2422   2444       PBH1 13.
FT   REPEAT     2455   2477       PBH1 14.
FT   REPEAT     2479   2501       PBH1 15.
FT   REPEAT     2512   2542       PBH1 16.
FT   REPEAT     2550   2582       PBH1 17.
FT   REPEAT     2589   2611       PBH1 18.
FT   REPEAT     2612   2633       PBH1 19.
FT   REPEAT     2638   2660       PBH1 20.
SQ   SEQUENCE   2894 AA;  322361 MW;  8B64F786537A16DF CRC64;

  Query Match             4.0%;  Score 271;  DB 1;  Length 2894;
  Best Local Similarity   19.5%;  Pred. No. 0.00037;
  Matches  315;  Conservative  224;  Mismatches  535;  Indels  542;  Gaps   80;

Qy          7 NFNYSDPVDNKNILYLDTHLNTLAN-----EPEKAFRITGNIWVIP-DRFSRNSNPNLNK 60
```



Search completed: September 10, 2004, 03:35:45
Job time : 23.7519 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 59.0697 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-60
Perfect score: 6759
Sequence: 1 MPITINNFNYSDPVDNKNIL.....NYASLLESTSTHWGFVPVSE 1291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6759	100.0	1291	2	Q93HT3	Q93ht3 clostridium
2	5072.5	75.0	1280	2	Q9LBS7	Q9lbs7 clostridium
3	5064.5	74.9	1280	2	Q841S3	Q841s3 clostridium
4	5049.5	74.7	1280	2	Q45849	Q45849 clostridium
5	4309	63.8	1285	2	Q45967	Q45967 clostridium
6	4307	63.7	1285	2	Q9LBR1	Q9lbr1 clostridium
7	3338.5	49.4	1275	12	Q9QTG7	Q9qtg7 clostridium
8	1935	28.6	1291	2	Q8GR96	Q8gr96 clostridium
9	1909	28.2	1291	2	Q08077	Q08077 clostridium
10	1907	28.2	1291	2	Q933K0	Q933k0 clostridium
11	1904	28.2	1291	2	Q9ZAJ8	Q9zaj8 clostridium
12	1895	28.0	1291	2	Q93G71	Q93g71 clostridium
13	1836	27.2	1278	2	Q57236	Q57236 clostridium
14	1823.5	27.0	1251	2	Q9K395	Q9k395 clostridium
15	1822.5	27.0	1310	2	Q93N27	Q93n27 clostridium
16	1820	26.9	1252	2	Q8KZM3	Q8kzm3 clostridium
17	1820	26.9	1255	2	Q9FAR6	Q9far6 clostridium
18	1804.5	26.7	1268	2	Q45851	Q45851 clostridium
19	1743.5	25.8	1280	2	Q9ZAJ5	Q9zaj5 clostridium
20	752	11.1	361	2	Q45848	Q45848 clostridium
21	751	11.1	361	2	Q45846	Q45846 clostridium
22	693.5	10.3	1197	2	Q45888	Q45888 clostridium
23	683.5	10.1	1198	2	O06018	O06018 clostridium
24	670	9.9	1197	2	Q9ZAJ9	Q9zaj9 clostridium
25	670	9.9	1197	2	O33871	O33871 clostridium
26	668.5	9.9	1197	2	P71117	P71117 clostridium
27	658	9.7	451	2	Q9R631	Q9r631 clostridium
28	657	9.7	1165	2	Q45844	Q45844 clostridium
29	655.5	9.7	1198	2	O69277	O69277 clostridium
30	654.5	9.7	1198	2	Q45893	Q45893 clostridium
31	651	9.6	1165	2	Q45887	Q45887 clostridium
32	644	9.5	367	2	Q45862	Q45862 clostridium
33	642.5	9.5	1161	2	O69276	O69276 clostridium
34	641	9.5	367	2	Q45861	Q45861 clostridium
35	640.5	9.5	1161	2	Q45891	Q45891 clostridium
36	640.5	9.5	1193	2	Q45880	Q45880 clostridium
37	640	9.5	1196	2	Q45916	Q45916 clostridium
38	640	9.5	1196	9	Q38197	Q38197 clostridium
39	639	9.5	1160	2	O87710	O87710 clostridium
40	638	9.4	1196	2	Q93HT4	Q93ht4 clostridium
41	637	9.4	1196	9	Q9ZX77	Q9zx77 clostridium
42	636	9.4	1196	2	Q53550	Q53550 clostridium
43	635.5	9.4	1193	2	Q45914	Q45914 clostridium
44	630.5	9.3	1193	2	P71107	P71107 clostridium
45	625	9.2	1196	2	Q9LBR2	Q9lbr2 clostridium

ALIGNMENTS

RESULT 1
Q93HT3
ID Q93HT3 PRELIMINARY; PRT; 1291 AA.
AC Q93HT3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-Yoichi;
RX MEDLINE=21534265; PubMed=11676492;
RA Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,
RA Fujinaga Y., Oguma K., Ohyama T.;
RT "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in
RT Hemagglutination.";
RL Biochem. Biophys. Res. Commun. 288:650-657(2001).
DR EMBL; AB061780; BAB71749.1; -.
DR PIR; A43503; A43503.
DR PIR; S11291; A49777.
DR PIR; S46431; S46431.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  148869 MW;  4A21DB35B8743CF8 CRC64;

  Query Match           100.0%;  Score 6759;  DB 2;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 6.8e-278;
  Matches 1291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60

Qy         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120

Qy        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360

Qy        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480

Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID 960

Qy        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM 1020

Qy       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL 1080

Qy       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN 1140

Qy       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA 1200

Qy       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR 1260

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              ||||||||||||||||||||||||||||||
Db       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291


RESULT 2
Q9LBS7
ID   Q9LBS7      PRELIMINARY;      PRT;  1280 AA.
AC   Q9LBS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-6814;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Kawabe T.,
RA   Murakami F., Nakatsuka M., Ohyama T.;
RT   "Organization of Gene Encoding Components of the Botulinum Progenitor
RT   Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT   Chimeric Sequence in the Gene Encoding Each Component.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037166; BAA89713.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147757 MW;  DE124FFB6F68450B CRC64;

  Query Match           75.0%;  Score 5072.5;  DB 2;  Length 1280;
  Best Local Similarity  77.5%;  Pred. No. 1.4e-206;
  Matches 1002;  Conservative   94;  Mismatches  170;  Indels   27;  Gaps   10;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              |||||||||||||||||||||||||||||||||||| |||||||||||:|||||||
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRIIGNIWVIPDRFSRDSNPNLNK 60
```

```
Qy        61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
             ||||||||||||||||||||||:||||||||||||||||||||||||||||||:|||||||||
Db        61 PPRVTSPKSGYYDPNYLSTDSEKDTFLKEIIKLFKRINSREIGEELIYRLATDIPFPGNN 120

Qy       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy       181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
             ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db       181 AAQEGFGALSIISISPRFMLTYSNATNNVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy       241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
             ||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       241 IAIPNDQRISSVTSNIFYSQYKVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
             |||||||||||||||||||||||||||||||||||||||| |:|||| ||| ||||||||
Db       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVAVDRNKFAELYKELTQIFTE 360

Qy       361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA 420

Qy       421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||| |
Db       421 LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDIKTDIFLSK 480

Qy       481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
             ||||||||| |||||||||||||||||||||||||| |: ||::|||||||||||||||||
Db       481 DINEETEVIDYPDNVSVDQVILSKNTSEHGQLDLLYPIIEGESQVLPGENQVFYDNRTQN 540

Qy       541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
             |||||||||||||||||||||||| |||||||| ||||||| ||:||| |||||||||
Db       541 VDYLNSYYYLESQKLSDNVEDFTFTTSIEEALDNSGKVYTYFPKLADKVNTGVQGGLFLM 600

Qy       601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660

Qy       661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
             ||| |||||||||||||||||||||||||||||||||||||||||||:||||||| |||
Db       661 EAFQEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMIGTWLSRITTQF 720

Qy       721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
             |||||||||||||| ||| ||||||||||||||||||||||||:||||||||||
Db       721 NNISYQMYDSLNYQADAIKDKIDLEYKKYSGSDKENIKSQVENLKNSLDIKISEAMNNIN 780

Qy       781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
             |||||||||||||||||||||||:||  || :|||||||||||||||||:|||||| ||
Db       781 KFIRECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESF 840

Qy       841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
             :|||||||||||||||||||||||:||||||||||:|| |||||||||||  |||||
Db       841 ENTIPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQ 900

Qy       901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTII 959
             :| |:  |||| |||:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||
Db       901 VNTIYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTII 957

Qy       960 DSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMG 1018
             :|:| |||| : | :  : : |:      :|: | |  | :  || |||||||:|||:||
Db       958 NSIKQNSGWKLCIRNGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMG 1017

Qy      1019 NMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAK 1078
              ||:||||:|  : ::::|  :   ||| | |::  |        :|  :|||| ||:|
Db      1018 YMKLYINGELKQSERIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSK 1069

Qy      1079 ELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN 1138
             ||  :||||::      ||:|||||| |::: |||::| :|::||:   |  :|     :
Db      1070 ELSNEDININVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPD 1129

Qy      1139 NNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHST 1195
```

```
             :    |  | || :      :|:  || : | |  |:  | :    :|:||      |
Db      1130 RSKLYTGNPITIKSVSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSK 1189

Qy      1196 EDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRL 1254
               :||: |:    :      || |: : :   | |||| |    :|   :  |   :||
Db      1190 NCVYALKLQSNLGNYGIG-IFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF 1246

Qy      1255 GGDWYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287
                    |  | |   || : | |||: | |:
Db      1247 ------ENAYTP-VAVTNYETKLLSTSSFWKFI 1272


RESULT 3
Q841S3
ID   Q841S3      PRELIMINARY;      PRT;  1280 AA.
AC   Q841S3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2003;
RA   Lee M.S., Cheng M.C., Chen Y.P., Shiau J.R., Lin S.Y.;
RT   "An outbreak of botulism in black-faced spoonbill in Taiwan.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY251553; AAP06952.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1280 AA;  147642 MW;  34C1A2E573EE0AE7 CRC64;

  Query Match           74.9%;  Score 5064.5;  DB 2;  Length 1280;
  Best Local Similarity   77.5%;  Pred. No. 3.1e-206;
  Matches 1003;  Conservative   93;  Mismatches  169;  Indels   29;  Gaps   11;

Qy         1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
             |||||||||||||||||||||||||||||||||||||| |||||||||||:|||||||
Db         1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRIIGNIWVIPDRFSRDSNPNLNK 60

Qy        61 PPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN 120
             ||| |||||||||||||||:|||||||||||||||||||||||||||:|||||||||
Db        61 PPRXTSPKSGYYDPNYLSTDSEKDTFLKEIIKLFKRINSREIGEELIYRLATDIPFPGNN 120

Qy       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF 180

Qy       181 AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240
             ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db       181 AAQEGFGALSIISISPRFMLTYSNATNNVGEGRFSKSEFCMDPILILMHELNHAMHNLYG 240

Qy       241 IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300
             ||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       241 IAIPNDQRISSVTSNIFYSQYKVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI 300

Qy       301 AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE 360
```

```
                                                
```

```
Qy        481 DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN 540
Db        481 DINVETEVIDYPDNVSVDQVILSKNTSEHGQLDLLYPIIEGESQVLPGENQVFYDNRTQN 540

Qy        541 VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM 600
Db        541 VDYLNSYYYLESQKLSDNVEDFTFTTSIEEALDNSGKVYTYFPKLADKVNTGVQGGLFLM 600

Qy        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL 660
Db        601 WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRENFTEAFAVTGVTILL 660

Qy        661 EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF 720
Db        661 EAFQEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMIGTWLSRITTQF 720

Qy        721 NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN 780
Db        721 NNISYQMYDSLNYQADAIKDKIDLEYKKYSGSDKENIKSQVENLKNSLDIKISEAMNNIN 780

Qy        781 KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF 840
Db        781 KFIRECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESF 840

Qy        841 QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ 900
Db        841 ENTIPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQ 900

Qy        901 LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTII 959
Db        901 VNTIYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTII 957

Qy        960 DSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMG 1018
Db        958 NSIKQNSGWKLCIRNGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMG 1017

Qy       1019 NMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAK 1078
Db       1018 YMKLYINGELKQSERIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSK 1069

Qy       1079 ELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN 1138
Db       1070 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPD 1129

Qy       1139 NNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHST 1195
Db       1130 RSKLYTGNPITIKSVSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSK 1189

Qy       1196 EDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRL 1254
Db       1190 NCVYALKLQSNLGNYGIG-IFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF 1246

Qy       1255 GGDWYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287
Db       1247 ------ENAYTP-VAVTNYETKLLSTSSFWKFI 1272
```

RESULT 5
Q45967
ID Q45967 PRELIMINARY; PRT; 1285 AA.
AC Q45967;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE D;
RX MEDLINE=96283801; PubMed=8679691;
RA Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA Ogumad K.;
RT "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT types C and D organisms.";
RL Biochim. Biophys. Acta 1307:123-126(1996).
DR EMBL; D38442; BAA07477.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1285 AA; 147365 MW; EEDE98E4ECA6C413 CRC64;

Query Match 63.8%; Score 4309; DB 2; Length 1285;
Best Local Similarity 65.1%; Pred. No. 2.8e-174;
Matches 840; Conservative 157; Mismatches 280; Indels 14; Gaps 6;

```
Qy          5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRP 64

Qy         65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124
Db         65 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 124

Qy        125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184
Db        125 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 184

Qy        185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244
Db        185 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 244

Qy        245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304
Db        245 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGSDVEIIPQIERLQLREKALGHYKDIAKRL 304

Qy        305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364
Db        305 NNINKTIPSSWSSNIDKYKKIFSEKYNFDKDNTGNFLVNIDKFNSLYSDLTNVMSEVIYS 364

Qy        365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424
Db        365 SQYNVKNRTHYFSKHYLPVFANILDDNIYTIINGFNLTTKGFNIENSGQNIERNPALQKL 424

Qy        425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484
Db        425 SSESVVDLFTKVCL-----RLTRNSRDDSTCIQVKNNTLPYVADKDSISQEIFESQIITD 479

Qy        485 ETEVIYYPDNVSVDQVIL-SKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDY 543
Db        480 ETNVENYSDNFSLDESILDAKVPTNPEAVDPLLPNVNMEPLNVPGEEEVFYDDITKDVDY 539

Qy        544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603
Db        540 LNSYYYLEAQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 599

Qy        604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663
Db        600 EVVEDFTTNIMKKDTLDKISDVSAIIPYIGPALNIGNSALRGNFKQAFATAGVAFLLEGF 659
```

```
Qy       664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723
Db       660 PEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHI 719

Qy       724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783
Db       720 SYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 779

Qy       784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNT 843
Db       780 RECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESFENT 839

Qy       844 IPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNP 903
Db       840 IPFNIFSYTNNSLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNP 899

Qy       904 IFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVK 963
Db       900 IFPFDFKLGSSGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSNLPGYTIIDSVK 959

Qy       964 NNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIY 1023
Db       960 NNSGWSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFVTITTNMMGNMMIY 1019

Qy      1024 INGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGK 1083
Db      1020 INGKLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDK 1079

Qy      1084 DINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFN 1143
Db      1080 DINILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFN 1139

Qy      1144 EGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAIG 1202
Db      1140 EGYKIIIKRIRGNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYKPSRNLGTDLVPLG 1194

Qy      1203 LREQTKD-INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGD-WYR 1260
Db      1195 ALDQPMDEIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLGDDYWFN 1254

Qy      1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db      1255 HEYLIPVIKIEHYASLLESTSTHWVFVPASE 1285
```

RESULT 6
Q9LBR1
ID Q9LBR1 PRELIMINARY; PRT; 1285 AA.
AC Q9LBR1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D-4947;
RA Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT "Characterization of the Progenitor Toxin Components Produced by
RT Clostridium botulinum Type D Strain 4947.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB037920; BAA90661.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1285 AA; 147352 MW; B63AFA487D570680 CRC64;

Query Match 63.7%; Score 4307; DB 2; Length 1285;
Best Local Similarity 65.1%; Pred. No. 3.4e-174;
Matches 840; Conservative 157; Mismatches 280; Indels 14; Gaps 6;

```
Qy         5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
Db         5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRP 64

Qy        65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124
Db        65 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 124

Qy       125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184
Db       125 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 184

Qy       185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244
Db       185 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 244

Qy       245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304
Db       245 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGSDVEIIPQIERLQLREKALGHYKDIAKRL 304

Qy       305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364
Db       305 NNINKTIPSSWSSNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSEVVYS 364

Qy       365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424
Db       365 SQYNVKNRTHYFSKHYLPVFANILDDNIYTIINGFNLTTKGFNIENSGQNIERNPALQKL 424

Qy       425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484
Db       425 SSESVVDLFTKVCL-----RLTRNSRDDSTCIQVKNNTLPYVADKDSISQEIFESQIITD 479

Qy       485 ETEVIYYPDNVSVDQVIL-SKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDY 543
Db       480 ETNVENYSDNFSLDESILDAKVPTNPEAVDPLLPNVNMEPLNVPGEEEVFYDDITKDVDY 539

Qy       544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603
Db       540 LNSYYYLEAQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 599

Qy       604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663
Db       600 EVVEDFTTNIMKKDTLDKISDVSAIIPYIGPALNIGNSALRGNFKQAFATAGVAFLLEGF 659

Qy       664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723
Db       660 PEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTRFNHI 719

Qy       724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783
Db       720 SYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 779

Qy       784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNT 843
Db       780 RECSVTYLFKNMLPKVIDELNKFDLKTKTELINLIDSHNIILVGEVDRLKAKVNESFENT 839

Qy       844 IPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNP 903
```

```
                 ||||||||||||||||:||||||:||||||||||:||||:|||||||||  ||:|||||
Db        840 IPFNIFSYTNNSLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNP 899

Qy        904 IFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVK 963
              ||||||||||||:||||||||||||:||||||||||||||||||||||||||||||||||
Db        900 IFPFDFKLGSSGDDRGKVIVTQNENIVYNAMYESFSISFWIRINKWVSNLPGYTIIDSVK 959

Qy        964 NNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIY 1023
              ||||||||||||||||||||||:||| ||||||||  || |||||||||:| ||||||  ||
Db        960 NNSGWSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFVTITTNMMGNMMIY 1019

Qy       1024 INGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGK 1083
              ||||||||||||||||||||||||::||||:|||||||||||||||||||||||||||| |
Db       1020 INGKLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDK 1079

Qy       1084 DINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFN 1143
              ||||||||||||||||||||||||:|||||:|::|:||||      ||||||:||||||
Db       1080 DINILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFN 1139

Qy       1144 EGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAIG 1202
              |||||||||||||||||||| ::|||: ||:|| |:|      ||  : :   |: :|
Db       1140 EGYKIIIKRIRGNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYKPSRNLGTDLVPLG 1194

Qy       1203 LREQTKD-INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGD-WYR 1260
                :| .| |     | ||| |   |||||:| |:    |  || |||:| |:|| | |:
Db       1195 ALDQPMDEIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLGDDYWFN 1254

Qy       1261 HNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
              | ||:| :|  :|||||||||| ||| ||
Db       1255 HEYLIPVIKIEHYASLLESTSTHWVFVPASE 1285


RESULT 7
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; P:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match             49.4%;  Score 3338.5;  DB 12;  Length 1275;
  Best Local Similarity   52.9%;  Pred. No. 3e-133;
  Matches  682;  Conservative  200;  Mismatches  375;  Indels   33;  Gaps   13;

Qy          5 INNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRV 64
              : :|||||||:: :|||     | |    | ||| || |||||||||| ::||:|:||||
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRP 64

Qy         65 TSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPI 124
              ||    ||||:|||||  |||||| |||||||| |:||::||  |    || |:::||
Db         65 TSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPE 124

Qy        125 NTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQE 184
              :||||     :: |:  :  :|  |  | |||:| || ||:|   |       :  : |
Db        125 DTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFE 184

Qy        185 GFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIP 244
              ||| |||: ::| |:||:|: |::        || |||||:: ||||| |::| |||| ||
Db        185 GFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYGINIP 244

Qy        245 NDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRL 304
              :|: |    |  |:||    ::: |:| |||  :::||:  |    |||| :|: |||||
Db        245 SDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGHYKDIAKRL 304

Qy        305 NSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTEFNYA 364
              |:|    |||:   | :||:    || |  :::|   || :||  ||::|| : :|  |:
Db        305 NNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSEVVYS 364

Qy        365 KIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKV 424
              |||:||  | |  | || |||||||:| |::|||:     |:   |||: |||||:|:
Db        365 SQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPALQKL 424

Qy        425 NPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRKDINE 484
              : |::: |||| |       |   |   :  ||| ||:: |   :  :|| | | :
Db        425 SSESVVDLFTKVCL-----RLTKNSRDDSTCIKVKNNRLPYVADKDSISQEIFENKIITD 479

Qy        485 ETEVIYYPDNVSVDQVILSKNTSEHGQL-DLLYPSIDSESEILPGENQVFYDNRTQNVDY 543
              || |  | |  |:|: ||      : :: | | |::: |   |||| ||||: |: |||
Db        480 ETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFYDDITKYVDY 539

Qy        544 LNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWAN 603
              |||||||||||||:|||: | | |:|||  | |:||: |:|| ||| ||| |||
Db        540 LNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWAN 599

Qy        604 DVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAF 663
              :|||||||::|||||||||| |||||||||| ||  |||| :|||  ||  ||| |
Db        600 EVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFATAGVAFLLEGF 659

Qy        664 PEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNI 723
              |||||||| |  || :||| :|||||:||||||:|||||||:||:  |||||  ||||:|
Db        660 PEFTIPALGVFTFYSSIQERKEKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHI 719

Qy        724 SYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 783
              :|||||:||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 NYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFI 779

Qy        784 RECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNT 843
              |||||||||||||||||||:||  || :|||||||||||||||:|||||| ||:||
Db        780 RECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVDRLKAKVNESFENT 839

Qy        844 IPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNP 903
              :|||||||||||||||||||:||||||||||:|| ||||||||||    :||||
Db        840 MPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNT 899

Qy        904 IFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSV 962
              |:  |||| |||:   |:||  | ||:|:::||: |:||||:|:|  ::|    ||||:|:
Db        900 IYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSI 956

Qy        963 KNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMK 1021
              : |||| : | :  : : |:       :|: | |  | :  || |||||||:|||:|| ||
Db        957 EQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMK 1016

Qy       1022 IYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELD 1081
              :||||:|  : |:::|  :   ||| | |::  |        :|  :|||||  ||:|||
```

```
Db       1017 LYINGELKQSQKIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELS 1068

Qy       1082 GKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNND 1141

Db       1069 NEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSK 1128

Qy       1142 FNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDI 1198

Db       1129 LYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCV 1188

Qy       1199 YAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGD 1257

Db       1189 YALKLQSNLGNYGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS---- 1241

Qy       1258 WYRHNYLVPTVKQGNYASLLESTSTHWGFV 1287

Db       1242 -FKNAYTPVAVT--NYETKLLSTSSFWKFI 1268


RESULT 8
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match           28.6%;  Score 1935;  DB 2;  Length 1291;
  Best Local Similarity  34.4%;  Pred. No. 6.9e-74;
  Matches  470;  Conservative  242;  Mismatches  500;  Indels  154;  Gaps   38;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59

Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111

Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171

Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231

Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285

Db        233 IHVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIITPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345

Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404

Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463

Db        404 NMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVRAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES 523

Db        455 FFIADKNSFSDDLSKNERIEYDTQSNYIENRSSIDELILDTNLISKIEL----PSENTES 510

Qy        524 -----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572

Db        511 LTDFNVDVPVYEKQPAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDAL 563

Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628

Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688

Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIK 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748

Db        681 TIDNALTKRDEKWIDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808

Db        741 YSEKEKSNINIDFNDINSKLNEGINQAVDNINNFINECSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868

Db        801 TLKKNLLNYIDENKLYLIGSAEYEKSKVDKHLKTIIPFDLSMYTNNTLIIEIFNKYNSEI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928

Db        861 LNNIILNLRYRDNNLIDLSGYGANVEVYDGVELND--KNQFKLTSS--TNSEIRVTQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982

Db        917 IIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKISIRGNRIIWTLT 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041

Db        977 DINGKTKSVFFEYSIREDISDYINRWFFVTITNN-SDNAKIYINGKLESNIDIKDIGEVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKD 1100

Db       1036 ANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYLKD 1086

Qy       1101 YWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN-------DFNEGYKII 1149

Db       1087 FWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEKFI 1146

Qy       1150 IKRIRGNT---NDTRVRGGDILYFDMTINNK-----AYNLFMKNE--TMYADNHSTEDIY 1199

Db       1147 IRR-KSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLFLANIYDSNEFY 1205

Qy       1200 -AIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG-- 1256
```

```
Db       1206 KTIQIKE----------YDEQPT----YSCQLLFKK--DEESTDEIGLIGIHRFYESGIV 1249

Qy       1257 -----------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
                          ||      :  ||:  |   |       :| |:|  |
Db       1250 LKDYKNYFCISKWY-----LKEVKRKPYNPNL---GCNWQFIPKDE 1287


RESULT 9
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match           28.2%;  Score 1909;  DB 2;  Length 1291;
  Best Local Similarity   33.9%;  Pred. No. 8.7e-73;
  Matches  463;  Conservative  255;  Mismatches  493;  Indels  156;  Gaps   38;

Qy         1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
               ||:|||||||:||:|| ||: ::          |||:||  ||:||:|::    | : |
Db         1 MPVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy        60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
               |        ||        ||||:||:|:  |: | : :|||| || |: :||:|:  :
Db        61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFFQTLIKLFNRIKSKPLGEKLLEMII 113

Qy       112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
               ||: |:   |:  |: ::   :|:          :   |  ::|| ||   : : ||
Db       114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy       172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
               : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db       174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy       232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
               | :| ||||       :||::          |: |    ::  |:| |||    :|  |
Db       233 IHVLHGLYGIKVDDLPIVPNEKK--------FFMQSTDTIQAEELYTFGGQDPSIISPST 284

Qy       286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
               |   :| |  :| |  ||| :      |  |  |  || |    ||:|| :| |: :::
Db       285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy       346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
               | :||  |   ||| | |: | :: |  | |:   ||   |:|: :| |: ||||
Db       344 SFNKLYKSLMLGFTEINIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy       405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
               |:   : ||| : |  |  ::: |::     : | |::    :      | :  | | :|
Db       404 NMGKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKVPGI------CID--VDNENL 454

Qy       464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSES 523
               || | :    |:   : :   |:  |  ::  ::::||  :|    :::|  || ::||
Db       455 FFIADKNSFSDDLSKNERVEYNTQNNYIGNDFPINELIL--DTDLISKIEL--PSENTES 510

Qy       524 ----------EILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
                         |  |   :|| |  |        : || ||    |: | : | | ::||
Db       511 LTDFNVDVPVYEKQPAIKKVFTDENT-------IFQYLYSQTFPLNIRDISLTSSFDDAL 563

Qy       573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
               |:|||::|        ||||   |: |||  |   :|:||     :  |:|||:|:| |
Db       564 LVSSKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLI 620

Qy       629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
               :|||| |||: :   :|||  || : | :|||  ||  || :|  |:: |  :  :|:||
Db       621 VPYIGLALNVGDETAKGNFESAFEIAGSSILLEFIPELLIPVVGVFLLESYIDNKNKIIK 680

Qy       689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
               |||| | :|:::| | |  ::  ||| : |||  |   || :|||| |::  |  :|
Db       681 TIDNALTKRVEKWIDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNI 740

Qy       749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
               ||  :| ||      :: : |:  |::||:||| || ||||:|| | |:|  : :| :||
Db       741 YSEEEKSNININFNDINSKLNDGINQAMDNINDFINECSVSYLMKKMIPLAVKKLLDFDN 800

Qy       809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
               |  |:| || : : |:| |:  |:||:   :  |||:: :|:|  :|  | |:| : |
Db       801 TLKKNLLNYIDENKLYLIGSVEDEKSKVDKYLKTIIPFDLSTYSNIEILIKIFNKYNSEI 860

Qy       869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
               :: ||:|: | | |:| ||| |:|      |:||      ||| || :   |: ||||:|
Db       861 LNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSAD--SKIRVTQNQN 916

Qy       929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
               |::|||:  ||:|||||| |:    : |     ||||: :|||||| | |  | :::||
Db       917 IIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLI 976

Qy       983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
                    :|: | |:|  :   | |:|||||:||| : | |||||| |   : :|::  :
Db       977 DINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVI 1035

Qy      1042 FSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKD 1100
               :  |||::: |  |    :|:: | ||  :|:  :|  ::    |:  :||
Db      1036 VNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKD 1086

Qy      1101 YWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN-------DFNEGYKII 1149
               :||| | ||||||| |    | |:     ::  :|:  :: | |       :   | | |
Db      1087 FWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFI 1146

Qy      1150 IKRIRG--NTNDTRVRGGDILYFDMTINNK-----AYNLFMKNE-----TMYADNHSTED 1197
               |:|      : ||  ||  | :: |: ::::    ||  | : |     ::: :|  | |
Db      1147 IRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISD--SNEF 1204

Qy      1198 IYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG- 1256
                 | ::|          : ||     |    :||   : |:   |  || :||   |
Db      1205 YKTIEIKE----------YDEQPS----YSCQLLFKK--DEESTDDIGLIGIHRFYESGV 1248

Qy      1257 -----------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
                         ||      :  ||:  | | |       :| |:|  |
Db      1249 LRKKYKDYFCISKWY-----LKEVKRKPYKSNL---GCNWQFIPKDE 1287
```

```
RESULT 10
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             28.2%;  Score 1907;  DB 2;  Length 1291;
  Best Local Similarity   33.8%;  Pred. No. 1.1e-72;
  Matches  463;  Conservative  253;  Mismatches  492;  Indels  162;  Gaps   37;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||:|||||||:||:|| ||: ::          |||:||  ||:||:|::     | : |
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
              |         ||        ||||:||:|:  |: ||: :|||| || |: :||:|:  :
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
              ||: |:   |:  |: ::   :|:          :   |  ::|| ||   : : ||
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
              : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
              | :| ||||       :||::          |: |    ::  |:| |||    :|  |
Db        233 IHVLHGLYGIKVNDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIISPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
              |   :| |  :| |  ||| :     |  |  |  || |   ||:|| :| |: :::
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
              | :||  |   ||| | |: | :: |  | |:   ||   |:|| :| |: ||||
Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
              |:   : ||| : |  |  ::: |::     : | |::    :       | :  | | ||
Db        404 NMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVIL----------SKNTSEHGQL 512
              || | :    |:  : |   |:  |  :: |:::||          |:||
Db        455 FFIADKNSFSDDLSKNERIAYNTQNNYIENDFSINELILDTDLISKIELPSENTESLTDF 514

Qy        513 DLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
              ::  |    :    |  ::| | |         : || ||    :: | : | | ::||
Db        515 NVYVPVYKKQ----PAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDAL 563

Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
              | |||::|         ||||   |: |||   |   :| ||      :  |:|||:|:| |
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVNDFVIEANKSSTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
              :|||| |||: |   :|||  || : | :||||  ||  || :|||:: | :  :|:|||:
Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIE 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
              ||:: | :| ::| | |  ::  ||| : |||  |   || :|||||  |::  |  :|
Db        681 TINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
              ||  :: ||     :: : |:  |::|::||| || ||||:|| | |:|  :::| :||
Db        741 YSEKERSNINIDFNDVNSKLNEGINQAIDNINNFINECSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
              :  |:| || : : |:|  :  |:||:   : :|||:: :||||::| :| |:| ::|
Db        801 TLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKTSIPFDLSTYTNNTILIEIFNKYNSDI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
              :: ||:|: | | |:| ||| |:|     |:||      ||| ||     |: ||||:|
Db        861 LNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA--NSKIRVTQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
              |::|||:  ||:|||||| |:    : |     ||||: :|||||  | |  | :::||
Db        917 IIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNMIIWTLI 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
                  :|: | | |  :   | |:|||||:|||   | ||||||||   | :::: :
Db        977 DINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKLESHIDIRDIREVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN----MWIRDFYIFAKELDGKDINILFNSLQYTNV 1097
                :  | |:::          ||:    :|:: | ||  ||   :|  ::    |:
Db       1036 ANDEIIPKLD-----------GNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEY 1083

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN-------DFNEGY 1146
              :||:||| | ||||||| |     | |:      ::  :|:  :: | |        | |
Db       1084 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGE 1143

Qy       1147 KIIIKRIRGNT---NDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGL 1203
              | ||:| : |:   ||  ||  | :| |    |: : ::|                  |
Db       1144 KFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYM--------------YKYFK 1187

Qy       1204 REQTK----DINDNIIF----QIQPMNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRL 1254
              :|: |     |:|:  |     ||:  :    |: | :||   : |:   |  || :||
Db       1188 KEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYE 1245

Qy       1255 GG-------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
                |               ||     :  ||:  | | |     :| |:|  |
Db       1246 SGIVFKEYKDYFCISKWY-----LKEVKRKPYNSKL---GCNWQFIPKDE 1287


RESULT 11
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
```

```
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match             28.2%;  Score 1904;  DB 2;  Length 1291;
  Best Local Similarity   33.7%;  Pred. No. 1.4e-72;
  Matches  462;  Conservative  254;  Mismatches  492;  Indels  162;  Gaps   37;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||:|||||||:||:|| ||: ::           |||:|| ||:||:|::    | : |
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
              |        ||        ||||:||:|:  |: ||: :|||| || |: :||:|:  :
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
              ||: |:   |:  |: ::   :|:        :   | ::|| ||   : : ||
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
              : |: ||::|||| :  :   | ::  ::|   : |   |::  :  || |||||||
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
              | :| ||||       :||::          |: |    ::  |:| |||    :|  |
Db        233 IHVLHGLYGIKVNDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIISPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
              |   :| |  :| |  ||| :     |  |  |  || |    ||:|| :| |: :::
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
              | :||  |    ||| | |: | :: |  | |:    ||    |:||: :| |: ||||
Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRAAYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
              |:   : ||| : |  |  ::: |::     : | |::    :       | :  | | ||
Db        404 NMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVIL----------SKNTSEHGQL 512
              || | :    |:   : |   |:  |  :: |::::||          |:||
Db        455 FFIADKNSFSDDLSKNERIAYNTQNNYIENDFSINELILDTDLISKIELPSENTESLTDF 514

Qy        513 DLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
              ::  |    :    |   ::| |  |        : || ||    :: | : | | ::||
Db        515 NVYVPVYKKQ----PAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDAL 563

Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
              | |||::|       ||||   |: |||  |   :|:||     :  |:|||:|:| |
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
              :|||| |||: |   :|||  || : | :|||  ||  || :|||:: | :  :|:|||:
Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIE 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
              ||:: | :| ::| | |  ::  ||| : |||  |   || :|||||  |::  |  :|
Db        681 TINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
              ||  :: ||     :: : |:  |::|::||| || ||||:|| | |:|  :::| :||
Db        741 YSEKERSNINIDFNDVNSKLNEGINQAIDNINNFINECSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
              :  |:| || : : |:|  :  |:||:   : :|||:: :||||::| :| |:| ::|
Db        801 TLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKTSIPFDLSTYTNNTILIEIFNKYNSDI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
              :: ||:|: | | |:| ||| |:|     |:||      ||| ||     |: | ||:|
Db        861 LNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA--NSKIRVIQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
              |::|||:  ||:|||||| |:    : |    ||||: :||||| | |  | :::||
Db        917 IIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNMIIWTLI 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
                    :|: | | |  :   | |:|||||:|||   | ||||||||   | :::: :
Db        977 DINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKLESHIDIRDIREVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN----MWIRDFYIFAKELDGKDINILFNSLQYTNV 1097
                :  | |:::             ||:    :|:: | ||  ||   :|  ::    |:
Db       1036 ANDEIIFKLD------------GNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEY 1083

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN-------DFNEGY 1146
              :||:||| | ||||||| |    | |:     ::  :|:  :: | |        |  |
Db       1084 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGE 1143

Qy       1147 KIIIKRIRGNT---NDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGL 1203
              | ||:| : |:   ||  ||  | :| |    |: : ::|                |
Db       1144 KFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYM--------------YKYFK 1187

Qy       1204 REQTK----DINDNIIF----QIQPMNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRL 1254
              :|: |    |:|:  |    ||:  :    |: | :||   : |:   |  || :||
Db       1188 KEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYE 1245

Qy       1255 GG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
                |             ||      :  ||:  | | |      :| |:|  |
Db       1246 SGIVFKEYKDYFCISKWY-----LKEVKRKPYNSKL---GCNWQFIPKDE 1287


RESULT 12
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
```

```
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             28.0%;  Score 1895;  DB 2;  Length 1291;
  Best Local Similarity   33.6%;  Pred. No. 3.4e-72;
  Matches  461;  Conservative  252;  Mismatches  495;  Indels  162;  Gaps   37;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
Db          1 MSVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         60 KPPRVTSPKSG--------YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLS 111
Db         61 K-------SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMII 113

Qy        112 TDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETS 171
Db        114 NGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETI 173

Qy        172 TFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHEL 231
Db        174 DIGIQNH-FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHEL 232

Qy        232 NHAMHNLYGI------AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSA 285
Db        233 IHVLHGLYGIKVNDLPIVPNEKK--------FFMQSTDAIQAEELYTFGGQDPSIISPST 284

Qy        286 RKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRN 345
Db        285 DKSIYDKVLQNFRGIVDRLNKVLVC-ISDPNININIYKNKFKDKYKFVEDSEGKYSIDVE 343

Qy        346 KFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKS 404
Db        344 SFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDK 403

Qy        405 NLNVLFMGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDL 463
Db        404 NMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMC-KSVKAPGI------CID--VDNEDL 454

Qy        464 PFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVIL----------SKNTSEHGQL 512
Db        455 FFIADKNSFSDDLSKNERIAYNTQNNYIDNDFSINELILDTDLISKIELPSENTESLTDF 514

Qy        513 DLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEAL 572
Db        515 NVYVPEYKKQ----PAIKKIFTDENT-------IFQYLYSQTFPLDIRDISLTSSFDDAL 563

Qy        573 DNSAKVYTYFP----TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAI 628
Db        564 LFSNKVYSFFSMDYIKTANKV---VEAGLFAGWVKQIVDDFVIEANKSSTMDKIADISLI 620

Qy        629 IPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIK 688
Db        621 VPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIE 680

Qy        689 TIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK 748
Db        681 TINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYKYNI 740

Qy        749 YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDR 808
Db        741 YSEKERSNINIDFNDVNSKLNEGINQAIDNINNFINECSVSYLMKKMIPLAVEKLLDFDN 800

Qy        809 NTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNI 868
Db        801 TLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKTSIPFDLSTYTNNTILIEIFNKYNSDI 860

Qy        869 NDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNEN 928
Db        861 LNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA--NSKIRVIQNQN 916

Qy        929 IVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLK 982
Db        917 IIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNMIIWTLI 976

Qy        983 QNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGIN 1041
Db        977 DINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKLESHIDIRDIREVI 1035

Qy       1042 FSKTITFEINKIPDTGLITSDSDNIN----MWIRDFYIFAKELDGKDINILFNSLQYTNV 1097
Db       1036 ANDEIIFKLD------------GNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEY 1083

Qy       1098 VKDYWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN-------DFNEGY 1146
Db       1084 LKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGE 1143

Qy       1147 KIIIKRIRGNT---NDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGL 1203
Db       1144 KFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVY--------------IYKYFK 1187

Qy       1204 REQTK----DINDNIIF----QIQPMNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRL 1254
Db       1188 KEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYE 1245

Qy       1255 GG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 1291
Db       1246 SGIVFKEYKDYFCISKWY-----LKEVKRKPYNSKL---GCNWQFIPKDE 1287


RESULT 13
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
```

```
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           27.2%;  Score 1836;  DB 2;  Length 1278;
  Best Local Similarity  33.9%;  Pred. No. 1.1e-69;
  Matches  456;  Conservative  245;  Mismatches  509;  Indels  136;  Gaps   41;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
              ||:.||:|||:|||::  |||:       : :  ||| |  |:|:||:|: :  ::| : :
Db          1 MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD 60

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
               | . :  | |||||||:||::|| :||  |||||||||   || |:  :|   |: ||
Db         61 PPASLENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGN 120

Qy        120 NNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF---KLT 176
              :|||| |       ||:|:       ||: ||  :::: |   :|   | |::   ||
Db        121 EHTPINEFHPVTRTTSVNIKSSTN---VKS-SIILNLLVLGAGPDIF--ENSSYPVRKLM 174

Qy        177 NNTFA---AQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSE-FCMDPILILMHELN 232
              ::       : :|||:::|:: || :  |:    ||:  |  | :| |  || : | |||
Db        175 DSGGVYDPSNDGFGSINIVTFSPEYEYTF----NDISGGYNSSTESFIADPAISLAHELI 230

Qy        233 HAMHNLYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEE 291
              ||:| |||   :   :||   : :  ::  ::||  |   |||  :::|  : ::
Db        231 HALHGLYGARGVTYKETIKVKQAPLMIAEKPIRLE--EFLTFGGQDLNIITSAMKEKIYN 288

Qy        292 KALDYYRSIAKRLNSITTANPSSFNKY-IGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL 350
              |  |  || ||: : :| |    :| | |||     ||     : |  ||| ||| |:
Db        289 NLLANYEKIATRLSRVNSAPP----EYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEI 344

Qy        351 YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLF 410
              | :|   ||| : |  : |: |  |           |:|||::| :  ||||   || |
Db        345 YKKLYS-FTEIDLANKFKVKCRNTYFIKYGFLKVPNLLDDDIYTVSEGFNI--GNLAVNN 401

Qy        411 MGQNLSRNPALRKVNPE-NMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDI 469
              |||:  || :    |:  ::    |||   |  :        |  : | | :| |:
Db        402 RGQNIKLNPKIIDSIPDKGLVEKIVKFCKSVIPRKGTKAPPRLC--IRVNNRELFFVASE 459

Qy        470 SDV-KTDIFLRKDINEETEV-IYYPDNVSVDQVILSKNTSEHGQL----------DLLY 516
              |   : ||   |:|:: | :   | :|  :|:|||  |:    |:           |..
Db        460 SSYNENDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSETIPQISNQTLNTLVQDDSYV 517

Qy        517 PSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSA 576
              | ||      | :::        ||  || ::|| :||: :   : : | ||: ||   :
Db        518 PRYDSN-----GTSEI----EEHNVVDLNVFFYLHAQKVPEGETNISLTSSIDTALSEES 568

Qy        577 KVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPA 635
              :|||:| :   | :|  |   ||: | | |: ||||   :| | |||:|:| ::||:| |
Db        569 QVYTFFSSEFINTINKPVHAALFISWINQVIRDFTTEATQKSTFDKIADISLVVPYVGLA 628

Qy        636 LNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKV---QERNEIIKTIDN 692
              ||| | |:: || ||| : |  ||||  ||  || :  | | | :   : :|:|||  |:|
Db        629 LNIGNEVQKENFKEAFELLGAGILLEFVPELLIPTILVFTIKSFIGSSENKNKIIKAINN 688

Qy        693 CLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGS 752
              | :|  :||: | |::  ||:|| ||||    ||| :|  |  |||  |: :|  |:
Db        689 SLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTVIEYKYNNYTSD 748

Qy        753 DKENIKSQ--VENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNT 810
              ::  ::|:  : |::  |: |:| || || :|| | |: || | :    : :| |:|
Db        749 ERNRLESEYNINNIREELNKKVSLAMENIERFITESSIFYLMKLINEAKVSKLREYDEGV 808

Qy        811 KAKLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNIND 870
              |  |:: |  |  ||   | :|   | ::  |:||| : ||||: :|    |: :  | |
Db        809 KEYLLDYISEHRSILGNSVQELNDLVTSTLNNSIPFELSSYTNDKILILYFNKLYKKIKD 868

Qy        871 SKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIV 930
              : || ::   |  :| ||| : :|  ||| :       | : ||     :| : || :|:
Db        869 NSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSS--KPSEVNIAQNNDII 926

Qy        931 YNSMYESFSISFWIRINKWVS--NLPG-YTIIDSVK-NNSGWSIGIISNFLVFTLKQNED 986
              ||  |::|||||:|| |: :  ||   ||||| :: ||||| | :  | :::||:
Db        927 YNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQDTAG 986

Qy        987 SEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKT 1045
              : | : |:|    :   | ||| |||:||| :|| :|||||  |||   :  |  |: |
Db        987 NNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDN 1046

Qy       1046 ITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGND 1105
              | |:|    ||  :           || | :|  ||   :|  |::    :::||:|||
Db       1047 ILFKIVGCNDTRYV---------GIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNY 1097

Qy       1106 LRYNKEYYMVNIDYLNRYMYANSRQIVFNTRR---------NNNDFNEGYKIIIKRIRGN 1156
              | ||| ||::|:   :: :  ||  :  | :|         :|    | ::||::  |:
Db       1098 LLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRK-NGS 1156

Qy       1157 TN----DTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDI-- 1210
              |:    |  ||  |: | ::   :  | |        |||          | : :  | |
Db       1157 TDISNTDNFVRKNDLAYINVVDRDVEYRL-------YAD---------ISIAKPEKIIKL 1200

Qy       1211 -----NDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRH 1261
                  ::| : ||  |::      :  |::| || |       ||   |      :  || :
Db       1201 IRTSNSNNSLGQIIVMDSIGNNCTMNFQNN-NGGN------IGLLGFHSNNLVASSWYYN 1253

Qy       1262 NYLVPTVKQGNYASLLESTSTHWGFV 1287
              |   |  | :          | |:
Db       1254 NIRKNTSSNGCF----------WSFI 1269


RESULT 14
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
```

```
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             27.0%;  Score 1823.5;  DB 2;  Length 1251;
  Best Local Similarity   33.5%;  Pred. No. 3.5e-69;
  Matches  447;  Conservative  264;  Mismatches  482;  Indels  143;  Gaps   42;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK 60
              ||  ||:|||:|||::: |||:          |  |:| |  |||:||:|    : |
Db          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Qy         61 PPRVTSPKSG---YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFP 117
              ||  || |:|   |||||||  :| :|| ||| : |:| |||:   |  |:  ||    |:
Db         57 PP--TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYL 114

Qy        118 GNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTN 177
              ||:||| | |    | ::|::|   |:  :    : |:||| |   :: :  :|   | |
Db        115 GNDNTPDNQFHIG-DASAVEIKFSNGSQDI----LLPNVIIMGAEPDLFETNSSNISLRN 169

Qy        178 NTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHN 237
              |   :  |||:::|:: ||     ||   ||       | :||  || | |||||  |::|
Db        170 NYMPSNHGFGSIAIVTFSPE----YSFRFND-----NSMNEFIQDPALTLMHELIHSLHG 220

Qy        238 LYGI-AIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDY 296
              |||   |    ||:  : :  :    :|  |   |||  :::|  :              |
Db        221 LYGAKGITTKYTITQKQNPLITNIRGTNIE--EFLTFGGTDLNIITSAQSNDIYTNLLAD 278

Qy        297 YRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQ 356
              |: || :|: :  :||       :  ||      ||    ::||  :|| ||| ::: :|
Db        279 YKKIASKLSKVQVSNP-----LLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYS 333

Qy        357 IFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLS 416
              ||||: |  : |: |: |:         :|:|:|::|:|  |:||  :|| | | ||| :
Db        334 -FTEFDLATKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNAN 390

Qy        417 RNP-ALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTD 475
              ||  :  :    ::    :||   :  : :  |:: | |  : | :| |:   :    |
Db        391 LNPRIITPITGRGLVKKIIRFCKNIVSVKGI-RKSI-CIE--INNGELFFVASENSYNDD 446

Qy        476 IFLRKDINEETEVIYYPDNVS--------VDQVILSKNTSEHGQLDLLYPSIDSESEILP 527
                :||   |:    | |:         :|||||: |:             | :  |  |
Db        447 -----NINTPKEI---DDTVTSNNNYENDLDQVILNFNSES-------APGLSDEKLNLT 491

Qy        528 GENQVF---YDN------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKV 578
              :|  :   ||:          :|: || ::||::||: :   :   | ||: ||    |:
Db        492 IQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKI 551

Qy        579 YTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637
              ||:| :   | ||  ||  ||: |   |: ||||   :| |:|||:|:| ::|||| |||
Db        552 YTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALN 611
```

```
Qy        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYS---KVQERNEIIKTIDNCL 694
Db        612 IGNEAQKGNFKDALELLGAGILLEFVPELLIPTILVFTIKSFLGSSDNKNKVIKAINNAL 671

Qy        695 EQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDK 754
Db        672 KERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKTIIEFKYNSYTLEEK 731

Qy        755 ENIKS--QVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKA 812
Db        732 KELKNNYDIEQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKT 791

Qy        813 KLINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSK 872
Db        792 YLLDYIIQHGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSS 851

Qy        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYN 932
Db        852 VLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFNS--KPSEVNISQNDYIIYD 909

Qy        933 SMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDS 987
Db        910 NKYKNFSISFWVRIPNYDNKIVNINNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNARI 969

Qy        988 EQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTI 1046
Db        970 NQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGHLIDQKSILNLGNIHVSDNI 1029

Qy       1047 TFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDL 1106
Db       1030 LFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL 1080

Qy       1107 RYNKEYYMVNIDYLNRYM--YANSRQIVFNTRRN---NNDFNEGYKIIIKRIR-GNTNDT 1160
Db       1081 LYDKGYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNDSSTNDR 1140

Qy       1161 RVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQT--KDINDNIIFQI 1218
Db       1141 FVRKNDQVYINYISNSSSYSL-------YADTNTTD-------KEKTIKSSSSGNRFNQV 1186

Qy       1219 QPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYA 1274
Db       1187 VVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVASTWY------------YT 1226

Qy       1275 SLLESTSTH---WGFV 1287
Db       1227 HMRDHTNSNGCFWNFI 1242
```

```
RESULT 15
Q93N27
ID   Q93N27      PRELIMINARY;      PRT;  1310 AA.
AC   Q93N27;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tetanus toxin (Fragment).
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shumin Z., Dianliang L.;
RT   "Cloning and sequence analysis of tetanus toxin gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF389424; AAK72964.2; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER    1310   1310
SQ   SEQUENCE   1310 AA;  150316 MW;  9EADDC914418E450 CRC64;

  Query Match             27.0%;  Score 1822.5;  DB 2;  Length 1310;
  Best Local Similarity   32.7%;  Pred. No. 4e-69;
  Matches  447;  Conservative  249;  Mismatches  531;  Indels  141;  Gaps   31;

Qy          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNP-NLN 59
Db          2 IPITINNFRYSVPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 61

Qy         60 KPPRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGN 119
Db         62 PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 121

Qy        120 NNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKL---T 176
Db        122 SYSLLDKFDTNSNSVSFNLSEQDPSGATTKSAMLTSLIIFGPGPVLNKNEVRGIVLRVDN 181

Qy        177 NNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMH 236
Db        182 KNYFPCRDGFGSIMQMTFCPEYIPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLH 241

Qy        237 NLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDY 296
Db        242 GLYGMQVSSHEIIPS--KQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLND 299

Qy        297 YRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQ 356
Db        300 YKAIANKLSQVTSCNDPNID--IDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMY 357

Qy        357 IFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLFMGQNL 415
Db        358 GFTEIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNM 417

Qy        416 SRNP-ALRKVNPENMLYLFTKFCHKAID----GRSLYNKTLDCREL------LVKNTDLP 464
Db        418 RVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTAALTDLGGELCIKIKNEDLI 477

Qy        465 FIGDISDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESE 524
Db        478 FIAEKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIILDYNL----QSKITLPN-DRTTP 532

Qy        525 ILPG-------ENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAK 577
Db        533 VTKGIPYAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTK 592

Qy        578 VYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALN 637
Db        593 IYSYFPSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALN 652

Qy        638 ISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQR 697
Db        653 IVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKR 712

Qy        698 IKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENI 757
Db        713 YEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQI 772

Qy        758 KSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINL 817
```

```
Db        773 ADEINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINETKKQLLEFDTQSKNILMQY 832

Qy        818 IDSHN-IILVGEVDKLKAKVNNSFQNTIPF----NIFSYTNNSLLKDIINEYFNNINDSK 872

Db        833 IKANSKFIGITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVI------LKKST 886

Qy        873 ILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGKVIVTQ 925

Db        887 ILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSEVIVHK 938

Qy        926 NENIVYNSMYESFSISFWIRINKWVS-------NLPGYTIIDSVKN-----NSGWSIGII 973

Db        939 AMDIEYNDMFNNFTVSFWLRVPK-VSASHLEQYDTNEYSIISSMKKYSLSIGSGWSVSLK 997

Qy        974 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGKLIDT 1031

Db        998 GNNLIWTLKDSAGEVRQITFR-DLSDKFNAYLANKWVFITITNDRLSSANLYINGVLMGS 1056

Qy       1032 IKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNS 1091

Db       1057 AEITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEKLYTS 1108

Qy       1092 LQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR---------YMYANSRQIVFNTRRN--NN 1140

Db       1109 YLSITFLRDFWGNPLRYDTEYYLIPVAYSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYR 1168

Qy       1141 DFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLFMKNETMYAD 1191

Db       1169 RLYSGLKFIIKRYTPNNEIDSFVRSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRV 1228

Qy       1192 NHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYR 1251

Db       1229 GYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL---KLYDDKDASLGLVGTHN 1273

Qy       1252 FRLGGD----------WYRHNYLVPTVKQGNYASLLESTST-HWGFVP 1288

Db       1274 GQIGNDPNRDILIASNWY-------------FNHLKDKTLTCDWYFVP 1308
```

Search completed: September 10, 2004, 03:41:42
Job time : 64.0697 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 27.0694 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 462
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	462	100.0	462	2	AAW68397	Aaw68397 Clostridi
2	436	94.4	450	3	AAY77136	Aay77136 Synthetic
3	436	94.4	450	4	AAB04092	Aab04092 Botulism
4	436	94.4	1291	2	AAY05814	Aay05814 Non-toxic
5	52	11.3	399	3	AAY78982	Aay78982 C. botuli
6	37	8.0	451	2	AAW68398	Aaw68398 Clostridi
7	27	5.8	451	2	AAW68395	Aaw68395 Clostridi
8	27	5.8	452	2	AAW68396	Aaw68396 Clostridi
9	26	5.6	448	2	AAW68399	Aaw68399 Clostridi
10	26	5.6	462	2	AAR95009	Aar95009 Type A ne
11	26	5.6	462	2	AAW68390	Aaw68390 Clostridi
12	26	5.6	472	2	AAW68393	Aaw68393 Clostridi
13	26	5.6	472	2	AAW68394	Aaw68394 Clostridi
14	26	5.6	473	2	AAW68400	Aaw68400 Clostridi
15	23	5.0	393	4	AAG68149	Aag68149 Streptomy
16	23	5.0	393	4	AAB74844	Aab74844 S. griseu
17	23	5.0	423	2	AAW68098	Aaw68098 Chlamydom
18	22	4.8	22	4	AAB67674	Aab67674 Amino aci
19	22	4.8	22	5	AAM49518	Aam49518 His-tag p
20	22	4.8	26	7	ADD31745	Add31745 Purificat
21	22	4.8	53	2	AAW36986	Aaw36986 HFKT pept
22	22	4.8	54	2	AAW36990	Aaw36990 HFKT pept
23	22	4.8	91	2	AAY09127	Aay09127 Recombina
24	22	4.8	93	2	AAY09128	Aay09128 Recombina
25	22	4.8	195	3	AAB11377	Aab11377 B. stearo
26	22	4.8	269	5	AAM49520	Aam49520 B. mori 1
27	22	4.8	269	5	AAM49541	Aam49541 B. mori T
28	22	4.8	323	2	AAW71185	Aaw71185 FP505 pro
29	22	4.8	332	4	AAE02191	Aae02191 Staphyloc
30	22	4.8	364	3	AAB13011	Aab13011 Q-tagged
31	22	4.8	567	4	AAU00220	Aau00220 Bad-DTTR
32	21	4.5	21	2	AAW23653	Aaw23653 Recombina
33	21	4.5	21	2	AAW23650	Aaw23650 Recombina
34	21	4.5	21	2	AAW23647	Aaw23647 Recombina
35	21	4.5	194	5	AAM51139	Aam51139 HIV TAT t
36	20	4.3	411	4	AAU00219	Aau00219 Bcl-Xl-DT
37	17	3.7	24	2	AAW19779	Aaw19779 Polyhisti
38	17	3.7	24	2	AAY06647	Aay06647 Polyhisti
39	17	3.7	24	2	AAY27015	Aay27015 Met Gly H
40	17	3.7	24	2	AAY09370	Aay09370 12 amino
41	17	3.7	24	3	AAY81959	Aay81959 Enterokin
42	17	3.7	24	4	AAB47057	Aab47057 Peptide f
43	17	3.7	24	6	ABG73858	Abg73858 Plasmid p
44	17	3.7	24	6	ABU10203	Abu10203 Peptide f
45	17	3.7	128	6	ABB99729	Abb99729 Amino aci

ALIGNMENTS

RESULT 1
AAW68397
ID AAW68397 standard; protein; 462 AA.
XX
AC AAW68397;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type C1 toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW botulism; BotC.
XX
OS Clostridium botulinum; serotype C1 strain Stockholm.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30588.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 45; Page 339-341; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC express BotC soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for

```
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             100.0%;  Score 462;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 2
AAY77136
ID   AAY77136 standard; protein; 450 AA.
XX
AC   AAY77136;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype C; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87214.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 24; Page 41-42; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 450 AA;

  Query Match             94.4%;  Score 436;  DB 3;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         15 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 74

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 134

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 194

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 254

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 314

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 374

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 434

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db        435 LLESTSTHWGFVPVSE 450
```

RESULT 3
AAB04092
ID AAB04092 standard; protein; 450 AA.
XX
AC AAB04092;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype C).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54486.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 5b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 450 AA;

Query Match 94.4%; Score 436; DB 4; Length 450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         15 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 74

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 134

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 194

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 254

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 314

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 374

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 434

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db        435 LLESTSTHWGFVPVSE 450
```

RESULT 4
AAY05814
ID AAY05814 standard; protein; 1291 AA.
XX
AC AAY05814;
XX
DT 02-AUG-1999 (first entry)
XX
DE Non-toxic modified botulinum toxin rBoNT/C.
XX
KW Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery; mutant.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 229
FT /note= "His in native toxin"
FT Misc-difference 230
FT /note= "Glu in native toxin"
FT Misc-difference 233
FT /note= "His in native toxin"
XX
PN WO9920306-A1.
XX
PD 29-APR-1999.
XX
PF 16-OCT-1998; 98WO-US021897.
XX
PR 20-OCT-1997; 97US-00954302.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Simpson L, Kiyatkin N, Maksymowych A;
XX
DR WPI; 1999-302646/25.
DR N-PSDB; AAX25521.
XX
PT Modified toxin useful for systemic delivery of oral vaccines and
PT therapeutic agents.
XX
PS Example 1; Page 31-33; 37pp; English.
XX
CC The present sequence represents a modified serotype C botulinum toxin,
CC termed rBoNT/C, in which amino acids His-229, Glu-230 and His-233 of the
CC native sequence are substituted by Gly, Thr and Asn, respectively, i.e.
CC the zinc binding motif (see AAY05817) of the light chain holotoxin is

```
CC   modified, resulting in loss of endoprotease activity. DNA coding for the
CC   modified botulinum toxin (see AAX25521) was assembled from 3 separate
CC   toxin fragments using PCR and site-directed mutagenesis. The modified
CC   recombinant botulinum toxin maintains its ability to translocate from the
CC   gut into the general circulation but is non-toxic. It can be used as an
CC   oral vaccine for antigenic peptides including botulinum toxin (i.e. an
CC   oral vaccine for botulism) or for the oral delivery of other therapeutic
CC   agents to the general circulation
XX
SQ   Sequence 1291 AA;

  Query Match             94.4%;  Score 436;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 5
AAY78982
ID   AAY78982 standard; protein; 399 AA.
XX
AC   AAY78982;
XX
DT   20-JUN-2000  (first entry)
XX
DE   C. botulinum type D toxin amino acid sequence.
XX
KW   Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
KW   protect.
XX
OS   Clostridium botulinum.
XX
PN   WO200005252-A1.
XX
PD   03-FEB-2000.
XX
PF   20-JUL-1999;   99WO-IB001301.
XX
PR   22-JUL-1998;   98ZA-00006538.
XX
PA   (AGRI-) AGRIC RES COUNCIL.
XX
PI   De Bruyn EE,  Botha AD;
XX
DR   WPI; 2000-205375/18.
DR   N-PSDB; AAZ98630, AAZ98631.
XX
PT   Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin type
PT   D, useful in vaccines for protection against botulism, comprises at least
PT   one amino acid mutation not present in the wild type D neurotoxins.
XX
PS   Claim 3, 4; Page 54-57; 66pp; English.
XX
CC   This sequence represents the amino acid sequence of a synthetic non-toxic
CC   immunogenic derivative of Clostridium botulinum type D toxin (BoNT).
CC   Botulinum neurotoxin causes botulism poisoning in cattle and sheep, and
CC   usually results in the death of the affected or poisoned animal. The non-
CC   toxic immunogenic fragments of the C. botulinum neurotoxin are useful in
CC   vaccines to protect animals (e.g. humans, cattle, sheep, pigs) against
CC   BoNT type D poisoning. The non-toxic fragments can be produced relatively
CC   simply and inexpensively (specifically by fermentation techniques). As
CC   the fragments are not toxic the risk to production staff is reduced
XX
SQ   Sequence 399 AA;

  Query Match             11.3%;  Score 52;  DB 3;  Length 399;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-44;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        105 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         44 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 95


RESULT 6
AAW68398
ID   AAW68398 standard; protein; 451 AA.
XX
AC   AAW68398;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type D toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW   botulism; BotD.
XX
OS   Clostridium botulinum; serotype D.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30591.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 47; Page 352-353; 428pp; English.
XX
```

```
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC   (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC   BotD soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             8.0%;  Score 37;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-28;
  Matches   37;  Conservative    0;  Mismatches   .0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFN 37
              |||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFN 37


RESULT 7
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.8%;  Score 27;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27
              |||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27


RESULT 8
AAW68396
ID   AAW68396 standard; protein; 452 AA.
XX
AC   AAW68396;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain NCTC 11219.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;
```

Query Match 5.8%; Score 27; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 2.5e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27
||||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27

RESULT 9
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30593.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC express BotC soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX
SQ Sequence 448 AA;

Query Match 5.6%; Score 26; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 2.5e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 10
AAR95009
ID AAR95009 standard; protein; 462 AA.
XX
AC AAR95009;
XX
DT 07-JUL-1996 (first entry)
XX
DE Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /label= Polyhistidine_tag
FT Protein 22. .462
FT /label= C-fragment
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29246.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 7; Page 340-342; 434pp; English.
XX
CC pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC fragment C (see also AAR95008) of the Clostridium botulinum type A
CC neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC soluble protein and was purified by metal chelate affinity chromatography
CC to obtain a product free of endotoxin contamination that may be useful as
CC an immunogen in vaccine compsns
XX
SQ Sequence 462 AA;

Query Match 5.6%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.6e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 11
AAW68390
ID AAW68390 standard; protein; 462 AA.
XX

```
AC   AAW68390;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW   Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW   botulism.
XX
OS   Clostridium botulinum; serotype A.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal histidine tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30572, AAV30575.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 24; Page 265-267; 428pp; English.
XX
CC   This is the amino acid sequence of a histidine-tagged fragment C
CC   polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC   sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC   AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC   non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on a poly-histidine binding
CC   affinity resin. The invention relates to recombinant proteins derived
CC   from C. botulinum toxins. Methods are provided which allow the isolation
CC   of soluble recombinant proteins that are free of significant endotoxin
CC   contamination. Preferred hosts for production of recombinant proteins are
CC   E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC   immunogens for the production of vaccines and antitoxins that are useful
CC   in the treatment of humans and animals at risk of intoxication with
CC   clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             5.6%;  Score 26;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 12
AAW68393
ID   AAW68393 standard; protein; 472 AA.
XX
AC   AAW68393;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B strain Eklund 17B.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30580.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 300-302; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             5.6%;  Score 26;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 13
AAW68394
ID   AAW68394 standard; protein; 472 AA.
XX
AC   AAW68394;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
```

```
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30581.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 303-305; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             5.6%;  Score 26;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 14
AAW68400
ID   AAW68400 standard; protein; 473 AA.
XX
AC   AAW68400;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW   botulism; BotG.
XX
OS   Clostridium botulinum; serotype G strain 113/30.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30596.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC   sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC   express BotG soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match             5.6%;  Score 26;  DB 2;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 15
AAG68149
ID   AAG68149 standard; protein; 393 AA.
XX
AC   AAG68149;
XX
DT   24-JAN-2002  (first entry)
XX
DE   Streptomyces griseus RppA protein with a His tag SEQ ID NO:5.
XX
KW   Streptomyces griseus; RppA; pyran-2-one; drug; agricultural.
XX
OS   Streptomyces griseus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= His_tag
FT   Protein         22. .393
FT                   /label= RppA_protein
XX
PN   JP2001238677-A.
XX
PD   04-SEP-2001.
XX
PF   01-MAR-2000; 2000JP-00055782.
XX
PR   01-MAR-2000; 2000JP-00055782.
XX
PA   (MITR ) MITSUBISHI RAYON CO LTD.
XX
DR   WPI; 2001-613891/71.
XX
PT   Preparation of pyran-2-one derivative, used as intermediate and as
```

PT additive for thermosensible transfer paper, comprises contacting carbonyl
PT compound with microbe containing polyketide synthase gene.
XX
PS Example 1; Page 10-11; 11pp; Japanese.
XX
CC The present invention describes the preparation of pyran-2-one
CC derivatives. The preparation method comprises contacting carbonyl
CC compounds with microbes containing a polyketide synthase gene. Pyran-2-
CC one derivatives can be used as synthetic intermediates for drugs and
CC agricultural chemicals. They can also be used as additives for
CC thermosensible transfer paper. The present sequence represents the
CC Streptomyces griseus RppA protein with an N-terminal His tag, which is
CC used in an example from the present invention
XX
SQ Sequence 393 AA;

Query Match 5.0%; Score 23; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

Search completed: September 10, 2004, 05:06:00
Job time : 28.0694 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 7.57473 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 462
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	5.6	462	1	US-08-480-604A-26	Sequence 26, Appl
2	26	5.6	462	2	US-08-405-496A-26	Sequence 26, Appl
3	26	5.6	462	3	US-08-915-136-26	Sequence 26, Appl
4	26	5.6	462	4	US-09-084-517-26	Sequence 26, Appl
5	23	5.0	23	1	US-08-480-604A-24	Sequence 24, Appl
6	23	5.0	23	2	US-08-405-496A-24	Sequence 24, Appl
7	23	5.0	23	3	US-08-915-136-24	Sequence 24, Appl
8	23	5.0	23	4	US-09-084-517-24	Sequence 24, Appl
9	22	4.8	22	4	US-09-660-742-3	Sequence 3, Appli
10	22	4.8	53	2	US-08-651-818A-19	Sequence 19, Appl
11	22	4.8	53	3	US-09-184-826-19	Sequence 19, Appl
12	22	4.8	54	2	US-08-651-818A-23	Sequence 23, Appl
13	22	4.8	54	3	US-09-184-826-23	Sequence 23, Appl
14	22	4.8	323	1	US-08-591-989-5	Sequence 5, Appli
15	21	4.5	21	2	US-08-651-818A-21	Sequence 21, Appl
16	21	4.5	21	3	US-09-184-826-21	Sequence 21, Appl
17	17	3.7	24	3	US-08-584-031-8	Sequence 8, Appli
18	17	3.7	24	3	US-08-780-496-8	Sequence 8, Appli
19	17	3.7	24	4	US-08-928-069-9	Sequence 9, Appli
20	15	3.2	20	4	US-09-186-489-6	Sequence 6, Appli
21	15	3.2	20	4	US-10-043-665B-6	Sequence 6, Appli
22	12	2.6	16	5	PCT-US94-04361-56	Sequence 56, Appl
23	12	2.6	17	2	US-08-600-783-13	Sequence 13, Appl
24	11	2.4	20	4	US-09-674-677-34	Sequence 34, Appl
25	11	2.4	24	4	US-09-292-858B-27	Sequence 27, Appl
26	11	2.4	254	3	US-09-167-434-7	Sequence 7, Appli
27	11	2.4	254	3	US-08-853-755-7	Sequence 7, Appli
28	11	2.4	302	2	US-08-203-532F-4	Sequence 4, Appli
29	11	2.4	302	3	US-08-950-860-16	Sequence 16, Appl
30	11	2.4	302	3	US-09-078-465-4	Sequence 4, Appli
31	11	2.4	302	5	PCT-US95-01882A-4	Sequence 4, Appli
32	11	2.4	303	2	US-08-203-532F-2	Sequence 2, Appli
33	11	2.4	303	3	US-09-078-465-2	Sequence 2, Appli
34	11	2.4	303	5	PCT-US95-01882A-2	Sequence 2, Appli
35	11	2.4	363	4	US-09-328-352-4930	Sequence 4930, Ap
36	11	2.4	414	5	PCT-US92-06840-2	Sequence 2, Appli
37	10	2.2	25	4	US-09-721-154-14	Sequence 14, Appl
38	10	2.2	335	3	US-09-095-117-2	Sequence 2, Appli
39	10	2.2	335	4	US-09-850-260-2	Sequence 2, Appli
40	10	2.2	449	2	US-08-927-394-2	Sequence 2, Appli
41	10	2.2	480	1	US-07-882-292-2	Sequence 2, Appli
42	10	2.2	480	2	US-08-331-644-2	Sequence 2, Appli
43	10	2.2	480	5	PCT-US93-04102-2	Sequence 2, Appli
44	10	2.2	633	3	US-08-557-006C-43	Sequence 43, Appl
45	10	2.2	763	2	US-08-677-862-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-26

Query Match 5.6%; Score 26; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-26

Query Match 5.6%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-26

Query Match 5.6%; Score 26; DB 3; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-17;

```
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             5.6%;  Score 26;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 5
US-08-480-604A-24
; Sequence 24, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-24

  Query Match             5.0%;  Score 23;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-16;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 6
US-08-405-496A-24
```

```
; Sequence 24, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-24

  Query Match             5.0%;  Score 23;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-16;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 7
US-08-915-136-24
; Sequence 24, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-24

  Query Match             5.0%;  Score 23;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-16;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 8
US-09-084-517-24
; Sequence 24, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
```

```
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,517
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CARROLL, PETER G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  OPHD-01610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  23 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-084-517-24

  Query Match             5.0%;  Score 23;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-16;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 9
US-09-660-742-3
; Sequence 3, Application US/09660742
; Patent No. 6485977
; GENERAL INFORMATION:
;  APPLICANT:  Collmer, Alan
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  RECOMBINANT CONSTRUCTS AND TECHNIQUES FOR DELIVERING TO
;  TITLE OF INVENTION:  EUCARYOTIC CELLS BACTERIAL PROTEINS THAT ARE SECRETED
;  TITLE OF INVENTION:  VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE:  19603/4021
;  CURRENT APPLICATION NUMBER: US/09/660,742
;  CURRENT FILING DATE:  2000-09-13
;  PRIOR APPLICATION NUMBER:  60/153,507
;  PRIOR FILING DATE:  1999-09-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 22
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Polypeptide
US-09-660-742-3

  Query Match             4.8%;  Score 22;  DB 4;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 10
US-08-651-818A-19
; Sequence 19, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-19

  Query Match             4.8%;  Score 22;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 11
US-09-184-826-19
; Sequence 19, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
```

```
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-19

  Query Match             4.8%;  Score 22;  DB 3;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 12
US-08-651-818A-23
; Sequence 23, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-23

  Query Match             4.8%;  Score 22;  DB 2;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 13
US-09-184-826-23
; Sequence 23, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-23

  Query Match             4.8%;  Score 22;  DB 3;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 14
US-08-591-989-5
```

```
; Sequence 5, Application US/08591989
; Patent No. 5795721
;  GENERAL INFORMATION:
;    APPLICANT:  Ross S. Rabin, Sumedha Jayasena
;    APPLICANT:  and Larry Gold
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC
;    TITLE OF INVENTION:  ACID LIGANDS OF ICP4
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 East Prentice Avenue, Suite #200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.40 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX 49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-591-989-5

  Query Match             4.8%;  Score 22;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db         17 MGHHHHHHHHHHSSGHIEGRHM 38


RESULT 15
US-08-651-818A-21
; Sequence 21, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-21

  Query Match             4.5%;  Score 21;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRH 21
              |||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRH 21


Search completed: September 10, 2004, 05:17:09
Job time : 8.57473 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 23.3701 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 462
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	462	100.0	462	12	US-10-272-898-62	Sequence 62, Appl
2	462	100.0	462	15	US-10-354-774-62	Sequence 62, Appl
3	462	100.0	462	15	US-10-271-012-62	Sequence 62, Appl
4	462	100.0	462	16	US-10-729-122-62	Sequence 62, Appl
5	462	100.0	462	16	US-10-729-039-62	Sequence 62, Appl
6	436	94.4	450	10	US-09-910-186A-10	Sequence 10, Appl
7	436	94.4	1291	12	US-10-272-898-60	Sequence 60, Appl
8	436	94.4	1291	15	US-10-354-774-60	Sequence 60, Appl
9	436	94.4	1291	15	US-10-271-012-60	Sequence 60, Appl
10	436	94.4	1291	15	US-10-452-024-3	Sequence 3, Appli
11	436	94.4	1291	15	US-10-452-024-135	Sequence 135, App
12	436	94.4	1291	15	US-10-452-024-136	Sequence 136, App
13	436	94.4	1291	15	US-10-205-516-6	Sequence 6, Appli
14	436	94.4	1291	16	US-10-729-122-60	Sequence 60, Appl
15	436	94.4	1291	16	US-10-729-039-60	Sequence 60, Appl
16	436	94.4	1301	15	US-10-205-516-20	Sequence 20, Appl
17	52	11.3	1285	15	US-10-452-024-165	Sequence 165, App
18	52	11.3	1285	15	US-10-452-024-166	Sequence 166, App
19	37	8.0	451	12	US-10-272-898-68	Sequence 68, Appl
20	37	8.0	451	15	US-10-354-774-68	Sequence 68, Appl
21	37	8.0	451	15	US-10-271-012-68	Sequence 68, Appl
22	37	8.0	451	16	US-10-729-122-68	Sequence 68, Appl
23	37	8.0	451	16	US-10-729-039-68	Sequence 68, Appl
24	27	5.8	451	12	US-10-272-898-54	Sequence 54, Appl
25	27	5.8	451	15	US-10-354-774-54	Sequence 54, Appl
26	27	5.8	451	15	US-10-271-012-54	Sequence 54, Appl
27	27	5.8	451	16	US-10-729-122-54	Sequence 54, Appl
28	27	5.8	451	16	US-10-729-039-54	Sequence 54, Appl
29	27	5.8	452	12	US-10-272-898-56	Sequence 56, Appl
30	27	5.8	452	15	US-10-354-774-56	Sequence 56, Appl
31	27	5.8	452	15	US-10-271-012-56	Sequence 56, Appl
32	27	5.8	452	16	US-10-729-122-56	Sequence 56, Appl
33	27	5.8	452	16	US-10-729-039-56	Sequence 56, Appl
34	26	5.6	448	12	US-10-272-898-73	Sequence 73, Appl
35	26	5.6	448	15	US-10-354-774-73	Sequence 73, Appl
36	26	5.6	448	15	US-10-271-012-73	Sequence 73, Appl
37	26	5.6	448	16	US-10-729-122-73	Sequence 73, Appl
38	26	5.6	448	16	US-10-729-039-73	Sequence 73, Appl
39	26	5.6	462	12	US-10-272-898-26	Sequence 26, Appl
40	26	5.6	462	15	US-10-354-774-26	Sequence 26, Appl
41	26	5.6	462	15	US-10-271-012-26	Sequence 26, Appl
42	26	5.6	462	16	US-10-729-122-26	Sequence 26, Appl
43	26	5.6	462	16	US-10-729-039-26	Sequence 26, Appl
44	26	5.6	472	12	US-10-272-898-44	Sequence 44, Appl
45	26	5.6	472	12	US-10-272-898-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-62
; Sequence 62, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:

```
;           LENGTH: 462 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-272-898-62

  Query Match            100.0%;  Score 462;  DB 12;  Length 462;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 2
US-10-354-774-62
; Sequence 62, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 462 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-354-774-62

  Query Match            100.0%;  Score 462;  DB 15;  Length 462;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 3
US-10-271-012-62
; Sequence 62, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 62:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 462 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-271-012-62

  Query Match             100.0%;  Score 462;  DB 15;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 4
US-10-729-122-62
; Sequence 62, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 62:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 462 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-729-122-62

  Query Match             100.0%;  Score 462;  DB 16;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 5
US-10-729-039-62
```

```
; Sequence 62, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 462 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-729-039-62

  Query Match           100.0%;  Score 462;  DB 16;  Length 462;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 6
US-09-910-186A-10
; Sequence 10, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 450
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
;   OTHER INFORMATION: sequence
US-09-910-186A-10

  Query Match           94.4%;  Score 436;  DB 10;  Length 450;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         15 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 74

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 134

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 194

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 254

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 314

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        315 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 374

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 434

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db        435 LLESTSTHWGFVPVSE 450


RESULT 7
US-10-272-898-60
; Sequence 60, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-272-898-60

  Query Match            94.4%;  Score 436;  DB 12;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 8
US-10-354-774-60
; Sequence 60, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-354-774-60

  Query Match            94.4%;  Score 436;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915
```

```
Qy        87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy       147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy       207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy       267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy       327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy       387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy       447 LLESTSTHWGFVPVSE 462
             ||||||||||||||||
Db      1276 LLESTSTHWGFVPVSE 1291
```

```
RESULT 9
US-10-271-012-60
; Sequence 60, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-271-012-60
```

```
  Query Match           94.4%;  Score 436;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy        87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy       147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy       207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy       267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy       327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy       387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy       447 LLESTSTHWGFVPVSE 462
             ||||||||||||||||
Db      1276 LLESTSTHWGFVPVSE 1291
```

```
RESULT 10
US-10-452-024-3
; Sequence 3, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-3
```

```
  Query Match           94.4%;  Score 436;  DB 15;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy        87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy       147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
                 ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 11
US-10-452-024-135
; Sequence 135, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 135
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum phage
US-10-452-024-135

  Query Match             94.4%;  Score 436;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
                 ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 12
US-10-452-024-136
; Sequence 136, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 136
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-136

  Query Match             94.4%;  Score 436;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
                 ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 13
US-10-205-516-6
; Sequence 6, Application US/10205516
; Publication No. US20040018589A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-6

  Query Match             94.4%;  Score 436;  DB 15;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 14
US-10-729-122-60
; Sequence 60, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-122-60

  Query Match             94.4%;  Score 436;  DB 16;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 15
US-10-729-039-60
; Sequence 60, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;
```

```
;              STREET: 220 Montgomery Street, Suite 2200
;              CITY: San Francisco
;              STATE: California
;              COUNTRY: United States of America
;              ZIP: 94104
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/729,039
;              FILING DATE: 05-Dec-2003
;              CLASSIFICATION: 424
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US/08/704,159
;              FILING DATE: 1996/08/28
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Ingolia, Diane E.
;              REGISTRATION NUMBER: 40,027
;              REFERENCE/DOCKET NUMBER: OPHD-02304
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (415) 705-8410
;              TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 60:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1291 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-039-60

  Query Match            94.4%;  Score 436;  DB 16;  Length 1291;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


Search completed: September 10, 2004, 05:52:53
Job time : 30.3701 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 6.40036 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 462
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	436	94.4	1291	2	S46431	botulinum neurotox
2	436	94.4	1291	2	A49777	botulinum neurotox
3	52	11.3	1285	2	S70582	botulinum neurotox
4	11	2.4	302	2	A55641	homeotic protein G
5	11	2.4	303	2	A56837	homeotic protein M
6	11	2.4	303	2	B49122	homeobox protein M
7	11	2.4	303	2	A48130	growth arrest-spec
8	11	2.4	354	2	S39406	homeotic protein o
9	11	2.4	355	2	S35345	otx1 protein - mou
10	11	2.4	355	2	I56547	homeodomain protei
11	11	2.4	414	2	A40350	transcription repr
12	11	2.4	414	2	A48273	delta/YY1/NF-E1/UC
13	11	2.4	420	2	T39712	hypothetical prote
14	11	2.4	427	2	T42516	hypothetical prote
15	11	2.4	1180	2	S69205	stripe a/b protein
16	11	2.4	1276	2	S11455	botulinum neurotox
17	10	2.2	132	2	I39193	gene HOXA1 protein
18	10	2.2	133	2	B30242	stem cell protein
19	10	2.2	222	2	I39192	gene HOXA1 protein
20	10	2.2	236	2	S41512	Brn-3b protein - m
21	10	2.2	298	2	S41469	homeotic protein M
22	10	2.2	309	2	T29293	hypothetical prote
23	10	2.2	321	2	T02987	myb-related protei
24	10	2.2	326	2	D83483	probable metal tra
25	10	2.2	331	2	A30242	homeotic protein E
26	10	2.2	335	2	G01448	homeobox protein H
27	10	2.2	351	1	KGZQHL	histidine-rich gly
28	10	2.2	375	2	A46390	cAMP receptor subt
29	10	2.2	387	2	A47446	HNF-3/fork head fa
30	10	2.2	390	1	A38565	polycomb (Pc) prot
31	10	2.2	391	2	H86187	hypothetical prote
32	10	2.2	407	2	G84783	probable pectinest
33	10	2.2	410	2	I38502	gene Brn-3b protei
34	10	2.2	411	2	I58156	Brn-3.2 - mouse
35	10	2.2	430	2	S66671	neuron-derived rec
36	10	2.2	451	2	A55909	transforming prote
37	10	2.2	469	2	I37451	HBF-G2 (HFK-2) pro
38	10	2.2	476	2	A54743	transcription fact
39	10	2.2	480	2	JH0672	brain factor 1 pro
40	10	2.2	580	2	T46024	hypothetical prote
41	10	2.2	623	1	S33167	gene pointed prote
42	10	2.2	625	2	S71930	neuron-derived rec
43	10	2.2	628	2	JC2493	neuron derived orp
44	10	2.2	633	1	A26030	serine/threonine-s
45	10	2.2	657	2	A29454	knob-associated hi

ALIGNMENTS

RESULT 1
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 94.4%; Score 436; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy          87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy         147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy         207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy         267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy         327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy         387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275
```

```
Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291
```

RESULT 2
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   94.4%; Score 436; DB 2; Length 1291;
  Best Local Similarity   100.0%; Pred. No. 0;
  Matches 436; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155
```

```
Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291
```

RESULT 3
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membr
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match   11.3%; Score 52; DB 2; Length 1285;
  Best Local Similarity   100.0%; Pred. No. 3.5e-45;
  Matches   52; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy        105 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 981
```

RESULT 4
A55641
homeotic protein GAX - human
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 24-Sep-1999
C;Accession: A55641
R;LePage, D.F.; Altomare, D.A.; Testa, J.R.; Walsh, K.
Genomics 24, 535-540, 1994
A;Title: Molecular cloning and localization of the human GAX gene to 7p21.
A;Reference number: A55641; MUID:95229154; PMID:7713505
A;Accession: A55641
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <LEP>
A;Cross-references: GB:L36328; NID:g557549; PIDN:AAA58497.1; PID:g557550
C;Genetics:
A;Gene: GAX
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;186-242/Domain: homeobox homology <HOX>

  Query Match   2.4%; Score 11; DB 2; Length 302;
  Best Local Similarity   100.0%; Pred. No. 0.0043;
  Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          2 GHHHHHHHHHH 12
              ||||||||||||
```

```
Db         67 GHHHHHHHHHH 77


RESULT 5
A56837
homeotic protein MOX2 - human
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 24-Sep-1999
C;Accession: A56837; S52168
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.S.; Theodorakis, K.; Mankoo, B.; Pachnis, V.;
Genomics 26, 550-555, 1995
A;Title: Isolation of the human MOX2 homeobox gene and localization to chromosome 7p22.1
A;Reference number: A56837; MUID:95331791; PMID:7607679
A;Accession: A56837
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-303 <GRI>
A;Cross-references: EMBL:X82629; NID:g732790; PIDN:CAA57949.1; PID:g732791
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.; Theodorakis, K.; Maukoo, B.; Pachnis, V.; K
submitted to the EMBL Data Library, November 1994
A;Description: Isolation of the human MOX2 homeobox gene and localization to chromosome
A;Reference number: S52168
A;Accession: S52168
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-7,'R',9-110,'R',112-122,'Q',124-157,'V',159-303 <GR2>
A;Cross-references: EMBL:X82629
C;Genetics:
A;Gene: GDB:MOX2
A;Cross-references: GDB:120703; OMIM:155970
A;Map position: 3q13-3q13
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.4%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0043;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 6
B49122
homeobox protein Mox-2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Dec-1993 #sequence_revision 17-Nov-1995 #text_change 17-Nov-2000
C;Accession: B49122; S41779; S29902
R;Candia, A.F.; Hu, J.; Crosby, J.; Lalley, P.A.; Noden, D.; Nadeau, J.H.; Wright, C.V.
Development 116, 1123-1136, 1992
A;Title: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially e
A;Reference number: A49122; MUID:93201999; PMID:1363541
A;Accession: B49122
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN1>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
A;Experimental source: C57BL
A;Note: the complete translation is not shown
R;Candia, A.F.; Kovalik, J.P.; Wright, C.V.E.
Nucleic Acids Res. 21, 4982, 1993
A;Title: Amino acid sequence of Mox-2 and comparison to its Xenopus and rat homologs.
A;Reference number: S41469; MUID:94232829; PMID:7909944
A;Accession: S41779
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN2>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
R;Candia, A.
submitted to the EMBL Data Library, October 1992
A;Reference number: S29902
A;Accession: S29902
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 186-246 <CAN>
A;Cross-references: EMBL:Z16406
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.4%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0043;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 7
A48130
growth arrest-specific homeobox protein Cvx - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A48130; S31976
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
Mol. Cell. Biol. 13, 3722-3733, 1993
A;Title: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated d
A;Reference number: A48130; MUID:93268321; PMID:8098844
A;Accession: A48130
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-303 <GOR>
A;Cross-references: GB:Z17223; NID:g57951; PIDN:CAA78931.1; PID:g57952
A;Experimental source: aorta
A;Note: sequence extracted from NCBI backbone (NCBIN:132842, NCBIP:132843)
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
submitted to the EMBL Data Library, October 1992
A;Description: Cvx: A diverged homeodomain gene which is rapidly down-regulated followi
A;Reference number: S31976
A;Accession: S31976
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-182,'S',184-303 <GO2>
A;Cross-references: EMBL:Z17223
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.4%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0043;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 8
S39406
homeotic protein otx1 - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 15-Oct-1999
C;Accession: S39406
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.;
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoi
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S39406
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-354 <SIM>
C;Genetics:
A;Gene: otx1
```

C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.4%; Score 11; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.0049;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 9
S35345
otx1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000
C;Accession: S35345; S29087
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; E
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S35345
A;Molecule type: DNA
A;Residues: 1-355 <SIM>
R;Simeone, A.; Acampora, D.; Gulisano, M.; Stornaiuolo, A.; Boncinelli, E.
Nature 358, 687-690, 1992
A;Title: Nested expression domains of four homeobox genes in developing rostral brain.
A;Reference number: S29084; MUID:92357147; PMID:1353865
A;Accession: S29087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 31-98 <SI2>
A;Cross-references: EMBL:X68883; NID:g53540; PIDN:CAA48754.1; PID:g53541
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.4%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0049;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 10
I56547
homeodomain protein otx1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C;Accession: I56547
R;Frantz, G.D.; Weimann, J.M.; Levin, M.E.; McConnell, S.K.
J. Neurosci. 14, 5725-5740, 1994
A;Title: Otx1 and Otx2 define layers and regions in developing cerebral cortex and cereb
A;Reference number: I56547; MUID:95016961; PMID:7931541
A;Accession: I56547
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: GB:L32602; NID:g535739; PIDN:AAA53557.1; PID:g535740
C;Genetics:
A;Gene: OTX1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.4%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0049;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 11
A40350
transcription repressor protein YY1 - human
N;Alternate names: transcription repressor/activator protein NF-E1
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Oct-1999
C;Accession: A40350; S78494; S33712; A56419
R;Shi, Y.; Seto, E.; Chang, L.S.; Shenk, T.
Cell 67, 377-388, 1991
A;Title: Transcriptional repression by YY1, a human GLI-Krueppel-related protein, and re
A;Reference number: A40350; MUID:92005716; PMID:1655281
A;Accession: A40350
A;Molecule type: mRNA
A;Residues: 1-414 <SHI>
A;Cross-references: GB:M77698; NID:g186767; PIDN:AAA59467.1; PID:g186768
A;Experimental source: HeLa cells
A;Note: the authors translated the codon CGC for residue 371 as Lys, CGA for residue 375
R;Whitson, R.H.; Huang, T.; Dang, J.; Itakura, K.
submitted to the EMBL Data Library, July 1992
A;Description: Observed and predicted DNA binding of a zinc finger protein which recogni
A;Reference number: S78494
A;Accession: S78494
A;Molecule type: mRNA
A;Residues: 1-195,'G',197-414 <WHI>
A;Cross-references: EMBL:Z14077; NID:g38010; PIDN:CAA78455.1; PID:g38011
R;Park, K.; Atchison, M.L.
Proc. Natl. Acad. Sci. U.S.A. 88, 9804-9808, 1991
A;Title: Isolation of a candidate repressor/activator, NF-E1 (YY-1), delta, that binds t
A;Reference number: A56419; MUID:92052179; PMID:1946405
A;Accession: S33712
A;Molecule type: mRNA
A;Residues: 1-64,'R',66-195,'G',197-414 <PAR>
A;Cross-references: GB:M76541; NID:g189173; PIDN:AAA59926.1; PID:g189174
C;Keywords: DNA binding; transcription regulation; zinc finger
F;298-320/Region: zinc finger CCHH motif
F;327-347/Region: zinc finger CCHH motif
F;355-377/Region: zinc finger CCHH motif
F;385-407/Region: zinc finger CCHH motif

Query Match 2.4%; Score 11; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 69 GHHHHHHHHHH 79

RESULT 12
A48273
delta/YY1/NF-E1/UCRBP transcription factor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 01-Dec-2000
C;Accession: A48273; A42055; A56418
R;Safrany, G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 5559-5563, 1993
A;Title: Characterization of the mouse gene that encodes the delta/YY1/NF-E1/UCRBP trans
A;Reference number: A48273; MUID:93296177; PMID:8516301
A;Accession: A48273
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-414 <RES>
A;Cross-references: GB:L13968; NID:g293847; PIDN:AAA40477.1; PID:g293849
R;Flanagan, J.R.; Becker, K.G.; Ennist, D.L.; Gleason, S.L.; Driggers, P.H.; Levi, B.Z ,
Mol. Cell. Biol. 12, 38-44, 1992
A;Title: Cloning of a negative transcription factor that binds to the upstream conserved

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

A;Reference number: A42055; MUID:92107191; PMID:1309593
A;Accession: A42055
A;Molecule type: mRNA
A;Residues: 1-414 <FLA>
A;Cross-references: GB:M73963; NID:g202270; PIDN:AAA40522.1; PID:g202271
A;Note: sequence extracted from NCBI backbone (NCBIN:74641, NCBIP:74642)
R;Hariharan, N.; Kelley, D.E.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 88, 9799-9803, 1991
A;Title: delta, a transcription factor that binds to downstream elements in several poly
A;Reference number: A56418; MUID:92052178; PMID:1946404
A;Accession: A56418
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218,'S',220-374,'G',376-414 <HAR>
A;Cross-references: GB:M74590; NID:g192940; PIDN:AAA37521.1; PID:g192941
C;Genetics:
A;Introns: 227/1; 281/2; 301/3; 354/3
C;Keywords: transcription factor; zinc finger

  Query Match 2.4%; Score 11; DB 2; Length 414;
  Best Local Similarity 100.0%; Pred. No. 0.0057;
  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 70 GHHHHHHHHHH 80

RESULT 13
T39712
hypothetical protein SPBC17D11.01 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: hypothetical protein SPBC20F10.11
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Jan-2000
C;Accession: T39712; T39882
R;Beck, A.; Reinhardt, R.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21872
A;Accession: T39712
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-420 <BEC>
A;Cross-references: EMBL:AL031739; PIDN:CAA21072.1; GSPDB:GN00067; SPDB:SPBC17D11.01
A;Experimental source: strain 972h-; cosmid c17D11
R;Lyne, M.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21888
A;Accession: T39882
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-281 <LYN>
A;Cross-references: EMBL:AL021747; PIDN:CAA16851.1; GSPDB:GN00067
C;Genetics:
A;Gene: SPBC20F10.11; SPDB:SPBC17D11.01
A;Map position: 2
A;Introns: 10/1; 28/2
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match 2.4%; Score 11; DB 2; Length 420;
  Best Local Similarity 100.0%; Pred. No. 0.0057;
  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 298 GHHHHHHHHHH 308

RESULT 14
T42516
hypothetical protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T42516
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42516
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-427 <YOS>
A;Cross-references: EMBL:D89156; NID:g1749519; PIDN:BAA13818.1; PID:g1749520
A;Experimental source: strain PR745
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match 2.4%; Score 11; DB 2; Length 427;
  Best Local Similarity 100.0%; Pred. No. 0.0058;
  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 298 GHHHHHHHHHH 308

RESULT 15
S69205
stripe a/b protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S69205
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube g
A;Reference number: S69205; MUID:96203082; PMID:8612588
A;Accession: S69205
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1180 <FRO>
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789
C;Keywords: alternative splicing

  Query Match 2.4%; Score 11; DB 2; Length 1180;
  Best Local Similarity 100.0%; Pred. No. 0.015;
  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HHHHHHHHHHS 13
||||||||||
Db 641 HHHHHHHHHHS 651

Search completed: September 10, 2004, 05:14:53
Job time : 7.50036 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 3.99288 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 462
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	436	94.4	1290	1	BXC1_CLOBO	P18640 clostridium
2	11	2.4	303	1	MOX2_HUMAN	P50222 homo sapien
3	11	2.4	303	1	MOX2_MOUSE	P32443 mus musculu
4	11	2.4	303	1	MOX2_RAT	P39020 rattus norv
5	11	2.4	354	1	OTX1_HUMAN	P32242 homo sapien
6	11	2.4	355	1	OTX1_MOUSE	P80205 mus musculu
7	11	2.4	355	1	OTX1_RAT	Q63410 rattus norv
8	11	2.4	414	1	TYY1_HUMAN	P25490 homo sapien
9	11	2.4	414	1	TYY1_MOUSE	Q00899 mus musculu
10	11	2.4	420	1	YBE1_SCHPO	O42980 schizosacch
11	11	2.4	466	1	ZIC3_MOUSE	Q62521 mus musculu
12	11	2.4	467	1	ZIC3_HUMAN	O60481 homo sapien
13	11	2.4	611	1	SNF1_CANGA	Q00372 candida gla
14	11	2.4	619	1	SNF1_CANTR	O94168 candida tro
15	11	2.4	1276	1	BXD_CLOBO	P19321 clostridium
16	10	2.2	202	1	HAN1_CHICK	Q90691 gallus gall
17	10	2.2	204	1	HAN1_SHEEP	Q28555 ovis aries
18	10	2.2	298	1	MOX2_XENLA	P39021 xenopus lae
19	10	2.2	331	1	HXA1_MOUSE	P09022 mus musculu
20	10	2.2	335	1	HXA1_HUMAN	P49639 homo sapien
21	10	2.2	351	1	HRPX_PLALO	P04929 plasmodium
22	10	2.2	375	1	CAR2_DICDI	P34907 dictyosteli
23	10	2.2	387	1	QIN_AVIS3	P56260 avian sarco
24	10	2.2	390	1	PC_DROME	P26017 drosophila
25	10	2.2	410	1	PO42_HUMAN	Q12837 homo sapien
26	10	2.2	411	1	PO42_MOUSE	Q63934 mus musculu
27	10	2.2	430	1	NOR2_RAT	Q63516 rattus norv
28	10	2.2	449	1	GAT6_HUMAN	Q92908 homo sapien
29	10	2.2	451	1	FXGB_CHICK	Q90964 gallus gall
30	10	2.2	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
31	10	2.2	465	1	HNF6_MOUSE	O08755 mus musculu
32	10	2.2	465	1	HNF6_RAT	P70512 rattus norv
33	10	2.2	469	1	FXGA_HUMAN	P55316 homo sapien
34	10	2.2	477	1	FXGB_HUMAN	P55315 homo sapien
35	10	2.2	480	1	FXGB_RAT	Q00939 rattus norv
36	10	2.2	481	1	FXGB_MOUSE	Q60987 mus musculu
37	10	2.2	551	1	CBX4_MOUSE	O55187 mus musculu
38	10	2.2	558	1	CBX4_HUMAN	O00257 homo sapien
39	10	2.2	620	1	SNF1_CANAL	P52497 candida alb
40	10	2.2	623	1	PNT1_DROME	P51022 drosophila
41	10	2.2	626	1	NR43_HUMAN	Q92570 homo sapien
42	10	2.2	628	1	NR43_RAT	P51179 rattus norv
43	10	2.2	633	1	SNF1_YEAST	P06782 saccharomyc
44	10	2.2	657	1	KNOB_PLAFN	P06719 plasmodium
45	10	2.2	665	1	LAMA_XENLA	P11048 xenopus lae

ALIGNMENTS

```
RESULT 1
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
```

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             94.4%;  Score 436;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 914

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 974

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1034

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1094

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1154

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1214

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1274

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1275 LLESTSTHWGFVPVSE 1290


RESULT 2
MOX2_HUMAN
ID   MOX2_HUMAN     STANDARD;      PRT;   303 AA.
AC   P50222; Q9UPL6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE   homeobox).
GN   MEOX2 OR MOX2 OR GAX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95331791; PubMed=7607679;
RA   Grigoriou M., Kastrinaki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA   Pachnis V., Karagogeos D.;
RT   "Isolation of the human MOX2 homeobox gene and localization to
RT   chromosome 7p22.1-p21.3.";
RL   Genomics 26:550-555(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95229154; PubMed=7713505;
RA   Lepage D.F., Walsh K.;
RT   "Molecular cloning and localization of the human GAX gene to 7p21.";
RL   Genomics 24:535-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 230-303 FROM N.A.
RA   Cordes M., Lacy M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle (By
CC       similarity).
```

```
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Embryo and placenta.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82629; CAA57949.1; -.
DR   EMBL; L36328; AAA58497.1; -.
DR   EMBL; BC017021; AAH17021.1; -.
DR   EMBL; AC004452; AAC06184.1; -.
DR   PIR; A55641; A55641.
DR   PIR; A56837; A56837.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04005; -.
DR   Genew; HGNC:7014; MEOX2.
DR   MIM; 600535; -.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DNA_BIND    186    245       HOMEOBOX.
FT   CONFLICT     58     58       G -> D (IN REF. 2).
FT   CONFLICT     79     79       MISSING (IN REF. 2).
SQ   SEQUENCE   303 AA;  33457 MW;  809ADE0CD090023D CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0073;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 3
MOX2_MOUSE
ID   MOX2_MOUSE     STANDARD;      PRT;   303 AA.
AC   P32443;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201999; PubMed=1363541;
RA   Candia A.F., Hu J., Crosby J., Lalley P.A., Noden D., Nadeau J.H.,
RA   Wright C.V.E.;
RT   "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are
RT   differentially expressed during early mesodermal patterning in mouse
RT   embryos.";
RL   Development 116:1123-1136(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94232829; PubMed=7909944;
RA   Candia A.F., Kovalik J.-P., Wright C.V.E.;
RT   "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat
RT   homologs.";
RL   Nucleic Acids Res. 21:4982-4982(1993).
RN   [3]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=95349593; PubMed=7623821;
RA   Andres V., Fisher S., Wearsch P., Walsh K.;
RT   "Regulation of Gax homeobox gene transcription by a combination of
RT   positive factors including myocyte-specific enhancer factor 2.";
RL   Mol. Cell. Biol. 15:4272-4281(1995).
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-
CC       8.5 dpc it is found on the entire epithelium of the somite. At 9.5
CC       dpc its expression is restricted to the sclerotome. At 10.5 dpc it
CC       is found in sclerotomally derived cells including the vertebral
CC       and costal precursors.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16406; CAA78899.1; -.
DR   EMBL; S79168; -; NOT_ANNOTATED_CDS.
DR   PIR; B49122; B49122.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04048; -.
DR   MGD; MGI:103219; Meox2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       63     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33506 MW;  41BD05FC39AA4427 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0073;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 4
MOX2_RAT
ID   MOX2_RAT       STANDARD;      PRT;   303 AA.
AC   P39020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
```

```
DE   Homeobox protein MOX-2 (Growth arrest-specific homeobox).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93268321; PubMed=8098844;
RA   Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,
RA   Walsh K.;
RT   "Molecular cloning of a diverged homeobox gene that is rapidly down-
RT   regulated during the G0/G1 transition in vascular smooth muscle
RT   cells.";
RL   Mol. Cell. Biol. 13:3722-3733(1993).
RN   [2]
RP   REVISIONS.
RA   Walsh K.;
RL   Submitted (MAR-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Aorta and heart. Also detected in lung and
CC       kidney.
CC   -!- INDUCTION: Rapidly and transiently down-regulated during the
CC       transition from G0 to G1 induced by mitogen stimulation.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z17223; CAA78931.1; -.
DR   PIR; A48130; A48130.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04054; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       64     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33605 MW;  7776642AEFA3A2E8 CRC64;

  Query Match              2.4%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0073;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
OTX1_HUMAN
ID   OTX1_HUMAN      STANDARD;      PRT;   354 AA.
AC   P32242;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS):
CC       5'-TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; BC007621; AAH07621.1; -.
DR   PIR; S39406; S39406.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02079; -.
DR   Genew; HGNC:8521; OTX1.
DR   MIM; 600036; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
```

```
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    301       HIS-RICH.
SQ   SEQUENCE   354 AA;  37327 MW;  E32C1E47464BDDAA CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.0084;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 6
OTX1_MOUSE
ID   OTX1_MOUSE     STANDARD;      PRT;   355 AA.
AC   P80205;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1 OR OTX-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A.,.Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 31-98 FROM N.A.
RX   MEDLINE=92357147; PubMed=1353865;
RA   Simeone A., Acampora D., Gulisano M., Stornaiuolo A., Boncinelli E.;
RT   "Nested expression domains of four homeobox genes in developing
RT   rostral brain.";
RL   Nature 358:687-690(1992).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC057105; AAH57105.1; -.
DR   EMBL; BC058354; AAH58354.1; -.
DR   EMBL; X68883; CAA48754.1; -.
DR   PIR; S35345; S35345.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02080; -.
DR   MGD; MGI:97450; Otx1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
SQ   SEQUENCE   355 AA;  37531 MW;  DA5136D9D4FC948F CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0084;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 7
OTX1_RAT
ID   OTX1_RAT       STANDARD;      PRT;   355 AA.
AC   Q63410; Q64203;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95016961; PubMed=7931541;
RA   Frantz G.D., Weimann J.M., Levin M.E., McConnell S.K.;
RT   "Otx1 and Otx2 define layers and regions in developing cerebral
```

```
RT   cortex and cerebellum.";
RL   J. Neurosci. 14:5725-5740(1994).
RN   [2]
RP   SEQUENCE OF 176-323 FROM N.A.
RX   MEDLINE=96108898; PubMed=8613727;
RA   Robel L., Ding M., James A.J., Lin X., Simeone A., Leckman J.F.,
RA   Vaccarino F.M.;
RT   "Fibroblast growth factor 2 increases Otx2 expression in precursor
RT   cells from mammalian telencephalon.";
RL   J. Neurosci. 15:7879-7891(1995).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3' (By similarity). May play a role in the specification
CC       or differentiation of neurons in the deep layers of the cerebral
CC       cortex, and also in cerebellar regionalization during early
CC       development.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: Expressed throughout the forebrain and
CC       midbrain during development, and in addition is also seen in
CC       discrete spatial and temporal domains in the developing cerebral
CC       cortex and cerebellum. Confined to a subpopulation of neurons in
CC       layers 5 and 6 within the adult cerebral cortex and during
CC       development expression is high in the progenitors of these deep-
CC       layer cells. Expressed in the developing cerebellum in spatially
CC       restricted regions of the external granular layer.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32602; AAA53557.1; -.
DR   EMBL; S81924; -; NOT_ANNOTATED_CDS.
DR   PIR; I56547; I56547.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
FT   CONFLICT    219    219       T -> A (IN REF. 2).
FT   CONFLICT    229    229       R -> G (IN REF. 2).
SQ   SEQUENCE   355 AA;  37602 MW;  C875871723D0B876 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0084;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
TYY1_HUMAN
ID   TYY1_HUMAN     STANDARD;      PRT;   414 AA.
AC   P25490; Q14935;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1).
GN   YY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92005716; PubMed=1655281;
RA   Shi Y., Seto E., Chang L.-S., Shenk T.;
RT   "Transcriptional repression by YY1, a human GLI-Kruppel-related
RT   protein, and relief of repression by adenovirus E1A protein.";
RL   Cell 67:377-388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=92052179; PubMed=1946405;
RA   Park K., Atchison M.;
RT   "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
RT   that binds to the immunoglobulin kappa 3' enhancer and the
RT   immunoglobulin heavy-chain mu E1 site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Whitson R.H., Huang T., Dang J., Itakura K.;
RL   Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98152931; PubMed=9493912;
RA   McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
RA   Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
RT   "Targeting of the YY1 transcription factor to the nucleolus and the
RT   nuclear matrix in situ: the C-terminus is a principal determinant for
RT   nuclear trafficking.";
RL   J. Cell. Biochem. 68:500-510(1998).
RN   [6]
RP   INTERACTION WITH YAF2.
RX   MEDLINE=97169297; PubMed=9016636;
RA   Kalenik J.L., Chen D., Bradley M.E., Chen S.-J., Lee T.-C.;
RT   "Yeast two-hybrid cloning of a novel zinc finger protein that
RT   interacts with the multifunctional transcription factor YY1.";
RL   Nucleic Acids Res. 25:843-849(1997).
```

```
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
RX   MEDLINE=97098436; PubMed=8942976;
RA   Houbaviy H.B., Usheva A., Shenk T., Burley S.K.;
RT   "Cocrystal structure of YY1 bound to the adeno-associated virus P5
RT   initiator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 353-379.
RX   MEDLINE=98308000; PubMed=9642075;
RA   Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
RA   Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
RT   "Design, synthesis and structure of a zinc finger with an artificial
RT   beta-turn.";
RL   J. Mol. Biol. 279:973-986(1998).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
CC       EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
CC       PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77698; AAA59467.1; -.
DR   EMBL; M76541; AAA59926.1; -.
DR   EMBL; Z14077; CAA78455.1; -.
DR   EMBL; BC037308; AAH37308.1; -.
DR   PIR; A40350; A40350.
DR   PDB; 1UBD; 23-DEC-96.
DR   PDB; 1ZNM; 01-APR-98.
DR   TRANSFAC; T00915; -.
DR   Genew; HGNC:12856; YY1.
DR   MIM; 600013; -.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     69       GLY-RICH.
FT   DOMAIN       70     80       POLY-HIS.
FT   DOMAIN      159    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE 1.
FT   ZN_FING     325    347       C2H2-TYPE 2.
FT   ZN_FING     353    377       C2H2-TYPE 3.
FT   ZN_FING     383    407       C2H2-TYPE 4.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION.
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN.
FT   CONFLICT     65     65       H -> R (IN REF. 2).
FT   CONFLICT    196    196       G -> R (IN REF. 1).
FT   STRAND      296    297
FT   TURN        301    302
FT   STRAND      306    307
FT   HELIX       310    317
FT   HELIX       318    320
FT   STRAND      325    326
FT   TURN        328    330
FT   STRAND      333    334
FT   HELIX       337    343
FT   HELIX       344    346
FT   TURN        347    347
FT   STRAND      353    354
FT   TURN        358    359
FT   STRAND      363    364
FT   HELIX       367    378
FT   TURN        388    389
FT   TURN        397    400
FT   HELIX       401    407
SQ   SEQUENCE   414 AA;  44712 MW;  058C05A0AD2D04E6 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0096;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 9
TYY1_MOUSE
ID   TYY1_MOUSE     STANDARD;      PRT;   414 AA.
AC   Q00899;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1) (UCR-motif DNA-binding protein).
GN   YY1 OR UCRBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107191; PubMed=1309593;
RA   Flanagan J.R., Becker K.G., Ennist D.L., Gleason S.L., Driggers P.H.,
RA   Levi B.-Z., Appella E., Ozato K.;
RT   "Cloning of a negative transcription factor that binds to the
RT   upstream conserved region of Moloney murine leukemia virus.";
RL   Mol. Cell. Biol. 12:38-44(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93296177; PubMed=8516301;
RA   Safrany G., Perry R.P.;
RT   "Characterization of the mouse gene that encodes the delta/YY1/NF-
RT   E1/UCRBP transcription factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5559-5563(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052178; PubMed=1946404;
RA   Hariharan N., Kelley D.E., Perry R.P.;
RT   "Delta, a transcription factor that binds to downstream elements in
RT   several polymerase II promoters, is a functionally versatile zinc
RT   finger protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9799-9803(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. BINDS TO
CC       THE UPSTREAM CONSERVED REGION (UCR) (5'-CGCCATTTT-3') OF MOLONEY
CC       MURINE LEUKEMIA VIRUS (MULV).
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73963; AAA40522.1; -.
DR   EMBL; L13968; AAA40477.1; -.
DR   EMBL; L13969; AAA40477.1; JOINED.
DR   EMBL; L13965; AAA40477.1; JOINED.
DR   EMBL; L13966; AAA40477.1; JOINED.
DR   EMBL; L13967; AAA40477.1; JOINED.
DR   EMBL; M74590; AAA37521.1; -.
DR   EMBL; BC055899; AAH55899.1; -.
DR   PIR; A48273; A48273.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T00278; -.
DR   TRANSFAC; T00865; -.
DR   MGD; MGI:99150; Yy1.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003700; F:transcription factor activity; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     70       GLY-RICH.
FT   DOMAIN       71     82       POLY-HIS.
FT   DOMAIN      161    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    407       C2H2-TYPE.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION
FT                                (BY SIMILARITY).
FT
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION (BY SIMILARITY).
FT
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN (BY SIMILARITY).
FT
FT   CONFLICT    219    219       F -> S (IN REF. 3).
FT   CONFLICT    375    375       R -> G (IN REF. 3).
SQ   SEQUENCE   414 AA;  44717 MW;  C012378288E984F9 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0096;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 10
YBE1_SCHPO
ID   YBE1_SCHPO     STANDARD;      PRT;   420 AA.
AC   O42980; P78807;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative thiol protease C17D11.01 (EC 3.4.22.-).
GN   SPBC17D11.01 OR SPBC20F10.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
```

```
RL   DNA Res. 4:363-369(1997).
CC   -!- SIMILARITY: Belongs to peptidase family C48.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031739; CAA21072.1; -.
DR   EMBL; AL021747; CAA16851.1; -.
DR   EMBL; D89156; BAA13818.1; -.
DR   GeneDB_SPombe; SPBC17D11.01; -.
DR   InterPro; IPR003653; Peptidase_C48.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Thiol protease.
FT   DOMAIN      299    312       POLY-HIS.
FT   CONFLICT     57     57       L -> W (IN REF. 2).
FT   CONFLICT    420    420       F -> FGCYTCVV (IN REF. 2).
SQ   SEQUENCE   420 AA;  47257 MW;  2125D94D9A2D9E87 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.0097;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 11
ZIC3_MOUSE
ID   ZIC3_MOUSE     STANDARD;      PRT;   466 AA.
AC   Q62521;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=96132843; PubMed=8557628;
RA   Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA   Chapman V.M., Mikoshiba K.;
RT   "The mouse zic gene family. Homologues of the Drosophila pair-rule
RT   gene odd-paired.";
RL   J. Biol. Chem. 271:1043-1047(1996).
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: CNS. A high level expression is seen in the
CC       cerebellum.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D70849; BAA11116.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T04671; -.
DR   MGD; MGI:106676; Zic3.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     96       POLY-HIS.
FT   ZN_FING     294    321       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     327    351       C2H2-TYPE 2.
FT   ZN_FING     357    381       C2H2-TYPE 3.
FT   ZN_FING     387    409       C2H2-TYPE 4.
SQ   SEQUENCE   466 AA;  50644 MW;  089749DF726F069B CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 12
ZIC3_HUMAN
ID   ZIC3_HUMAN     STANDARD;      PRT;   467 AA.
AC   O60481;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HTX1 MET-323.
RX   MEDLINE=98016414; PubMed=9354794;
RA   Gebbia M., Ferrero G.B., Pilia G., Bassi M.T., Aylsworth A.S.,
RA   Penman-Splitt M., Bird L.M., Bamforth J.S., Burn J., Schlessiner D.,
RA   Nelson D.L., Casey B.;
RT   "X-linked situs abnormalities result from mutations in ZIC3.";
RL   Nat. Genet. 17:305-308(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in ZIC3 are the cause of X-Linked visceral
CC       heterotaxy (HTX1) [MIM:306955]. HTX1 is a disease characterized by
CC       congenital heart disease and alterations of visceral situs,
CC       including asplenia or polysplenia, symmetric liver, intestinal
CC       malrotation, and abnormal lung lobation.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028706; AAC05594.1; -.
DR   EMBL; AL035443; CAB41648.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:12874; ZIC3.
DR   MIM; 300265; -.
DR   MIM; 306955; -.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein; Disease mutation.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     97       POLY-HIS.
FT   ZN_FING     295    322       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     328    352       C2H2-TYPE 2.
FT   ZN_FING     358    382       C2H2-TYPE 3.
FT   ZN_FING     388    410       C2H2-TYPE 4.
FT   VARIANT     323    323       T -> M (in HTX1).
FT                                /FTId=VAR_007753.
SQ   SEQUENCE   467 AA;  50569 MW;  3150CF13C0679568 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 13
SNF1_CANGA
ID   SNF1_CANGA     STANDARD;      PRT;   611 AA.
AC   Q00372;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCCLS84;
RX   MEDLINE=97101049; PubMed=8945576;
RA   Petter R., Kwon-Chung K.J.;
RT   "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT   pathogenic yeast Candida glabrata.";
RL   Infect. Immun. 64:5269-5273(1996).
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78130; AAB48642.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       6     17       POLY-HIS.
FT   DOMAIN      39    290       PROTEIN KINASE.
FT   NP_BIND     45     53       ATP (BY SIMILARITY).
FT   BINDING     68     68       ATP (BY SIMILARITY).
FT   ACT_SITE   161    161       BY SIMILARITY.
FT   MOD_RES    194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  70049 MW;  89E17812A4900CD0 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db          8 HHHHHHHHHHS 18


RESULT 14
SNF1_CANTR
ID   SNF1_CANTR     STANDARD;      PRT;   619 AA.
AC   O94168;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanai T., Ogawa K., Ueda M., Tanaka A.;
RT   "Genetic evaluation of the function of SNF1 in Candida tropicalis.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB024535; BAA75889.1; -.
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
```

```
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       17     29       POLY-HIS.
FT   DOMAIN       52    303       PROTEIN KINASE.
FT   NP_BIND      58     66       ATP (BY SIMILARITY).
FT   BINDING      81     81       ATP (BY SIMILARITY).
FT   ACT_SITE    174    174       BY SIMILARITY.
FT   MOD_RES     207    207       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  70323 MW;  0FCF1FC3DCE706D7 CRC64;

  Query Match           2.4%;  Score 11;  DB 1;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db         20 HHHHHHHHHHS 30


RESULT 15
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN        1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN      443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL      229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   230    230       BY SIMILARITY.
FT   METAL      233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID   437    450       INTERCHAIN (PROBABLE).
FT   VARIANT     15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT     17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT    452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT    457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT    457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT    462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT    489    489       K -> N (IN STRAIN CB16).
FT   VARIANT    644    644       N -> K (IN STRAIN CB16).
FT   VARIANT   1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match           2.4%;  Score 11;  DB 1;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFN 37
              ||||||||||
Db        852 LLKDIINEYFN 862


Search completed: September 10, 2004, 05:07:21
Job time : 4.99288 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 18.79 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 462
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	436	94.4	1291	2	Q93HT3	Q93ht3 clostridium
2	52	11.3	1285	2	Q45967	Q45967 clostridium
3	52	11.3	1285	2	Q9LBR1	Q9lbr1 clostridium
4	12	2.6	226	11	Q9QZD2	Q9qzd2 rattus norv
5	12	2.6	516	13	O42186	O42186 brachydanio
6	12	2.6	517	13	Q90YL6	Q90yl6 brachydanio
7	12	2.6	569	13	Q802D2	Q802d2 brachydanio
8	12	2.6	679	13	Q803K9	Q803k9 brachydanio
9	12	2.6	706	5	Q8SXG1	Q8sxg1 drosophila
10	12	2.6	861	5	Q9VRT8	Q9vrt8 drosophila
11	11	2.4	163	5	Q7YTA3	Q7yta3 glomeris ma
12	11	2.4	225	13	Q90XZ3	Q90xz3 erpetoichth
13	11	2.4	230	4	O75263	O75263 homo sapien
14	11	2.4	243	4	Q9P2R2	Q9p2r2 homo sapien
15	11	2.4	243	6	Q9N2G2	Q9n2g2 gorilla gor
16	11	2.4	243	6	Q9N2G3	Q9n2g3 pan troglod
17	11	2.4	243	6	Q9N2G1	Q9n2g1 pongo pygma
18	11	2.4	302	13	Q90YH7	Q90yh7 gallus gall
19	11	2.4	303	6	Q95JA6	Q95ja6 sus scrofa
20	11	2.4	303	11	Q99M23	Q99m23 mus musculu
21	11	2.4	355	11	Q91ZJ3	Q91zj3 mus musculu
22	11	2.4	358	13	Q9DG32	Q9dg32 xenopus lae
23	11	2.4	411	11	Q8CHJ4	Q8chj4 rattus norv
24	11	2.4	441	5	Q8T1W3	Q8t1w3 dictyosteli
25	11	2.4	449	10	Q8L4M9	Q8l4m9 oryza sativ
26	11	2.4	456	11	Q8BSB3	Q8bsb3 mus musculu
27	11	2.4	467	13	Q9YGC6	Q9ygc6 xenopus lae
28	11	2.4	511	10	Q8GXZ5	Q8gxz5 arabidopsis
29	11	2.4	906	5	Q8IN91	Q8in91 drosophila
30	11	2.4	1122	5	Q8SXD4	Q8sxd4 drosophila
31	11	2.4	1147	5	Q95PH8	Q95ph8 dictyosteli
32	11	2.4	1180	5	Q24163	Q24163 drosophila
33	11	2.4	1186	5	Q9VED3	Q9ved3 drosophila
34	11	2.4	1275	12	Q9QTG7	Q9qtg7 clostridium
35	11	2.4	1280	2	Q45849	Q45849 clostridium
36	11	2.4	1280	2	Q9LBS7	Q9lbs7 clostridium
37	11	2.4	1280	2	Q841S3	Q841s3 clostridium
38	11	2.4	1283	5	Q95RH4	Q95rh4 drosophila
39	11	2.4	1911	5	Q9W4M7	Q9w4m7 drosophila
40	10	2.2	53	2	O55244	O55244 escherichia
41	10	2.2	233	4	Q8TC53	Q8tc53 homo sapien
42	10	2.2	233	5	Q8MP30	Q8mp30 dictyosteli
43	10	2.2	236	11	Q9QWE1	Q9qwe1 mus sp. cla
44	10	2.2	239	5	Q7Z1G6	Q7z1g6 trypanosoma
45	10	2.2	244	13	Q8JHA3	Q8jha3 gallus gall

ALIGNMENTS

RESULT 1
Q93HT3
ID Q93HT3 PRELIMINARY; PRT; 1291 AA.
AC Q93HT3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-Yoichi;
RX MEDLINE=21534265; PubMed=11676492;
RA Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,
RA Fujinaga Y., Oguma K., Ohyama T.;
RT "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in
RT Hemagglutination.";
RL Biochem. Biophys. Res. Commun. 288:650-657(2001).
DR EMBL; AB061780; BAB71749.1; -.
DR PIR; A43503; A43503.
DR PIR; S11291; A49777.
DR PIR; S46431; S46431.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  148869 MW;  4A21DB35B8743CF8 CRC64;

  Query Match             94.4%;  Score 436;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 915

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISN 975

Qy        147 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 FLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKVK 1035

Qy        207 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 ELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYT 1095

Qy        267 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 NVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRG 1155

Qy        327 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 NTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNII 1215

Qy        387 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 FQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYAS 1275

Qy        447 LLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1276 LLESTSTHWGFVPVSE 1291


RESULT 2
Q45967
ID   Q45967      PRELIMINARY;      PRT;  1285 AA.
AC   Q45967;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE D;
RX   MEDLINE=96283801; PubMed=8679691;
RA   Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA   Ogumad K.;
RT   "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT   types C and D organisms.";
RL   Biochim. Biophys. Acta 1307:123-126(1996).
DR   EMBL; D38442; BAA07477.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1285 AA;  147365 MW;  EEDE98E4ECA6C413 CRC64;

  Query Match             11.3%;  Score 52;  DB 2;  Length 1285;
  Best Local Similarity   100.0%;  Pred. No. 1e-45;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        105 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 981


RESULT 3
Q9LBR1
ID   Q9LBR1      PRELIMINARY;      PRT;  1285 AA.
AC   Q9LBR1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-4947;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA   Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT   "Characterization of the Progenitor Toxin Components Produced by
RT   Clostridium botulinum Type D Strain 4947.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037920; BAA90661.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1285 AA;  147352 MW;  B63AFA487D570680 CRC64;

  Query Match             11.3%;  Score 52;  DB 2;  Length 1285;
  Best Local Similarity   100.0%;  Pred. No. 1e-45;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        105 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 MYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNE 981


RESULT 4
Q9QZD2
ID   Q9QZD2      PRELIMINARY;      PRT;   226 AA.
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
```

```
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
SQ   SEQUENCE   226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

  Query Match             2.6%;  Score 12;  DB 11;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 0.00074;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db         17 HHHHHHHHHHSS 28


RESULT 5
O42186
ID   O42186      PRELIMINARY;      PRT;   516 AA.
AC   O42186;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FTZ-F1 homolog.
GN   NR5A2 OR ZFF1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322094; PubMed=9178748;
RA   Liu D., Le Drean Y., Ekker M., Xiong F., Hew C.L.;
RT   "Teleost FTZ-F1 homolog and its splicing variant determine the
RT   expression of the salmon gonadotropin IIbeta subunit gene.";
RL   Mol. Endocrinol. 11:877-890(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AF014926; AAC60274.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04065; -.
DR   ZFIN; ZDB-GENE-990415-79; nr5a2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   516 AA;  58805 MW;  500DA09E116D5FA5 CRC64;

  Query Match             2.6%;  Score 12;  DB 13;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 6
Q90YL6
ID   Q90YL6      PRELIMINARY;      PRT;   517 AA.
AC   Q90YL6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nuclear receptor FF1aE.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chai C., Tan J.H., Gao W., Teh H.L., Chan W.K.;
RT   "Full length zebrafish ff1aE cDNA.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AY034061; AAK54449.1; -.
DR   HSSP; P19793; 1LBD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   517 AA;  58761 MW;  CF48D08AB7A75F29 CRC64;

  Query Match             2.6%;  Score 12;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 7
Q802D2
ID   Q802D2      PRELIMINARY;      PRT;   569 AA.
AC   Q802D2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048886; AAH48886.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
KW   Cyclin.
SQ   SEQUENCE   569 AA;  64117 MW;  812A7332C06EF4C2 CRC64;

  Query Match             2.6%;  Score 12;  DB 13;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 0.0017;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        486 HHHHHHHHHHSS 497


RESULT 8
Q803K9
ID   Q803K9      PRELIMINARY;      PRT;   679 AA.
AC   Q803K9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T1 (Fragment).
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044435; AAH44435.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
KW   Cyclin.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  74887 MW;  AB5667D59439659F CRC64;

  Query Match             2.6%;  Score 12;  DB 13;  Length 679;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        488 HHHHHHHHHHSS 499


RESULT 9
Q8SXG1
ID   Q8SXG1      PRELIMINARY;      PRT;   706 AA.
AC   Q8SXG1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RH12258p.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AY089654; AAL90392.1; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   706 AA;  79683 MW;  73C0F8199DF60E5F CRC64;

  Query Match             2.6%;  Score 12;  DB 5;  Length 706;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        370 GHHHHHHHHHHS 381


RESULT 10
Q9VRT8
ID   Q9VRT8      PRELIMINARY;      PRT;   861 AA.
AC   Q9VRT8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG6619 protein.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
```

```
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AE003564; AAF50702.2; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   861 AA;  96809 MW;  8530ED5D92671A41 CRC64;

  Query Match             2.6%;  Score 12;  DB 5;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 0.0025;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        525 GHHHHHHHHHHS 536


RESULT 11
Q7YTA3
ID   Q7YTA3      PRELIMINARY;      PRT;   163 AA.
AC   Q7YTA3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Distal-less protein (Fragment).
GN   DLL.
OS   Glomeris marginata.
OC   Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Pentazonia;
OC   Glomerida; Glomeridae; Glomeris.
OX   NCBI_TaxID=62006;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22767835; PubMed=12885558;
RA   Prpic N.M., Tautz D.;
RT   "The expression of the proximodistal axis patterning genes Distal-less
RT   and dach shund in the appendages of Glomeris marginata (Myriapoda:
RT   Diplopoda) suggests a special role for these genes in patterning the
RT   head appendages.";
RL   Dev. Biol. 260:97-112(2003).
DR   EMBL; AJ551276; CAD82905.1; -.
FT   NON_TER       1      1
FT   NON_TER     163    163
SQ   SEQUENCE   163 AA;  17588 MW;  87B8BCD9F5421697 CRC64;

  Query Match             2.4%;  Score 11;  DB 5;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 0.0063;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHH 11
              ||||||||||
Db         16 MGHHHHHHHHH 26


RESULT 12
Q90XZ3
ID   Q90XZ3      PRELIMINARY;      PRT;   225 AA.
AC   Q90XZ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Homeodomain protein Otx1 (Fragment).
OS   Erpetoichthys calabaricus (Rope fish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Polypteriformes; Polypteridae; Erpetoichthys.
OX   NCBI_TaxID=27687;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21396394; PubMed=11504847;
RA   Germot A., Lecointre G., Plouhinec J.L., Le Mentec C., Girardot F.,
RA   Mazan S.;
RT   "Structural evolution of otx genes in Craniates.";
RL   Mol. Biol. Evol. 18:1668-1678(2001).
DR   EMBL; AF321406; AAL01673.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     225    225
SQ   SEQUENCE   225 AA;  23475 MW;  0AE78B8DB48B4941 CRC64;

  Query Match             2.4%;  Score 11;  DB 13;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 0.0085;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        172 GHHHHHHHHHH 182


RESULT 13
O75263
ID   O75263      PRELIMINARY;      PRT;   230 AA.
AC   O75263;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   WUGSC:H_DJ0620P06.1 protein (Fragment).
GN   WUGSC:H_DJ0620P06.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Andrews S., Lacy M., Dubbelde C.;
RT   "The sequence of Homo sapiens PAC clone RP4-620P6.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005550; AAC33152.1; -.
DR   HSSP; P14653; 1B72.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
FT   NON_TER     230    230
SQ   SEQUENCE   230 AA;  25476 MW;  81D627D0CFEC00B0 CRC64;

  Query Match             2.4%;  Score 11;  DB 4;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 0.0087;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 14
Q9P2R2
ID   Q9P2R2      PRELIMINARY;      PRT;   243 AA.
AC   Q9P2R2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Homeobox protein OTX1 (Fragment).
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037501; BAA90421.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     243    243
SQ   SEQUENCE   243 AA;  24428 MW;  01FA60D616BBAF65 CRC64;

  Query Match             2.4%;  Score 11;  DB 4;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 0.0091;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        190 GHHHHHHHHHH 200


RESULT 15
Q9N2G2
ID   Q9N2G2      PRELIMINARY;      PRT;   243 AA.
AC   Q9N2G2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Homeobox protein OTX1 (Fragment).
GN   OTX1.
OS   Gorilla gorilla (gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9593;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=gorilla-U1;
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037503; BAA90423.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     243    243
SQ   SEQUENCE   243 AA;  24428 MW;  01FA60D616BBAF65 CRC64;

  Query Match             2.4%;  Score 11;  DB 6;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 0.0091;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        190 GHHHHHHHHHH 200


Search completed: September 10, 2004, 05:12:56
Job time : 19.79 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 30.4751 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 2479
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2479	100.0	462	2	AAW68397	Aaw68397 Clostridi
2	2317	93.5	450	3	AAY77136	Aay77136 Synthetic
3	2317	93.5	450	4	AAB04092	Aab04092 Botulism
4	2317	93.5	1291	2	AAY05814	Aay05814 Non-toxic
5	1589	64.1	399	3	AAY78982	Aay78982 C. botuli
6	963.5	38.9	451	2	AAW68398	Aaw68398 Clostridi
7	801.5	32.3	451	4	AAB04093	Aab04093 Botulism
8	725.5	29.3	473	2	AAW68400	Aaw68400 Clostridi
9	718.5	29.0	462	2	AAR95009	Aar95009 Type A ne
10	718.5	29.0	462	2	AAW68390	Aaw68390 Clostridi
11	715	28.8	472	2	AAW68393	Aaw68393 Clostridi
12	711.5	28.7	452	2	AAW68396	Aaw68396 Clostridi
13	709	28.6	451	2	AAW68395	Aaw68395 Clostridi
14	700	28.2	472	2	AAW68394	Aaw68394 Clostridi
15	671.5	27.1	448	2	AAW68399	Aaw68399 Clostridi
16	596.5	24.1	445	2	AAW68391	Aaw68391 Clostridi
17	573.5	23.1	449	3	AAY77139	Aay77139 Synthetic
18	573.5	23.1	449	4	AAB04167	Aab04167 Botulism
19	564.5	22.8	438	2	AAR95008	Aar95008 Type A ne
20	564.5	22.8	438	2	AAW68389	Aaw68389 Clostridi
21	564.5	22.8	438	3	AAY77134	Aay77134 Synthetic
22	559.5	22.6	432	3	AAY77142	Aay77142 Native bo
23	557.5	22.5	437	4	AAB04088	Aab04088 Botulism
24	557.5	22.5	837	3	AAY77140	Aay77140 Native bo
25	557.5	22.5	1067	3	AAY93307	Aay93307 A mangane
26	557.5	22.5	1092	3	AAY93310	Aay93310 A mangane
27	557.5	22.5	1296	2	AAR95010	Aar95010 C. botuli
28	556.5	22.4	434	4	AAB04089	Aab04089 Botulism
29	556.5	22.4	435	4	AAB04090	Aab04090 Botulism
30	553.5	22.3	419	4	AAB04095	Aab04095 Botulism
31	553.5	22.3	449	3	AAY77137	Aay77137 Synthetic
32	553.5	22.3	449	4	AAB04094	Aab04094 Botulism
33	548.5	22.1	1295	5	AAU99339	Aau99339 Clostridi
34	548	22.1	1070	3	AAY93308	Aay93308 A mangane
35	548	22.1	1095	3	AAY93311	Aay93311 A mangane
36	548	22.1	1291	2	AAW68392	Aaw68392 Clostridi
37	542	21.9	440	3	AAY77135	Aay77135 Synthetic
38	542	21.9	440	4	AAB04091	Aab04091 Botulism
39	540	21.8	848	4	AAB04082	Aab04082 Botulism
40	539.5	21.8	847	4	AAB04081	Aab04081 Botulism
41	538	21.7	439	4	AAB04085	Aab04085 Botulism
42	529.5	21.4	862	4	AAE07890	Aae07890 Modified
43	529.5	21.4	887	4	AAE07892	Aae07892 Modified
44	529.5	21.4	979	6	AAE35713	Aae35713 BoNT/F-Hc
45	529.5	21.4	1192	6	AAE35711	Aae35711 BoNT/F-Hc

ALIGNMENTS

RESULT 1
AAW68397
ID AAW68397 standard; protein; 462 AA.
XX
AC AAW68397;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type C1 toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW botulism; BotC.
XX
OS Clostridium botulinum; serotype C1 strain Stockholm.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30588.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 45; Page 339-341; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC express BotC soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for

```
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             100.0%;  Score 2479;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-169;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 2
AAY77136
ID   AAY77136 standard; protein; 450 AA.
XX
AC   AAY77136;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype C; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87214.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 24; Page 41-42; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 450 AA;

  Query Match             93.5%;  Score 2317;  DB 3;  Length 450;
  Best Local Similarity   99.8%;  Pred. No. 1.7e-157;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 73

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 133

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 193

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 253

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 313

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 373

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 433

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db        434 SLLESTSTHWGFVPVSE 450
```

```
RESULT 3
AAB04092
ID   AAB04092 standard; protein; 450 AA.
XX
AC   AAB04092;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype C).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54486.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 5b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 450 AA;

  Query Match             93.5%;  Score 2317;  DB 4;  Length 450;
  Best Local Similarity   99.8%;  Pred. No. 1.7e-157;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 73

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 133

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 193

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 253

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 313

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 373

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 433

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db        434 SLLESTSTHWGFVPVSE 450


RESULT 4
AAY05814
ID   AAY05814 standard; protein; 1291 AA.
XX
AC   AAY05814;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Non-toxic modified botulinum toxin rBoNT/C.
XX
KW   Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery; mutant.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 229
FT                   /note= "His in native toxin"
FT   Misc-difference 230
FT                   /note= "Glu in native toxin"
FT   Misc-difference 233
FT                   /note= "His in native toxin"
XX
PN   WO9920306-A1.
XX
PD   29-APR-1999.
XX
PF   16-OCT-1998;   98WO-US021897.
XX
PR   20-OCT-1997;   97US-00954302.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Simpson L,  Kiyatkin N,  Maksymowych A;
XX
DR   WPI; 1999-302646/25.
DR   N-PSDB; AAX25521.
XX
PT   Modified toxin useful for systemic delivery of oral vaccines and
PT   therapeutic agents.
XX
PS   Example 1; Page 31-33; 37pp; English.
XX
CC   The present sequence represents a modified serotype C botulinum toxin,
CC   termed rBoNT/C, in which amino acids His-229, Glu-230 and His-233 of the
CC   native sequence are substituted by Gly, Thr and Asn, respectively, i.e.
CC   the zinc binding motif (see AAY05817) of the light chain holotoxin is
```

```
CC   modified, resulting in loss of endoprotease activity. DNA coding for the
CC   modified botulinum toxin (see AAX25521) was assembled from 3 separate
CC   toxin fragments using PCR and site-directed mutagenesis. The modified
CC   recombinant botulinum toxin maintains its ability to translocate from the
CC   gut into the general circulation but is non-toxic. It can be used as an
CC   oral vaccine for antigenic peptides including botulinum toxin (i.e. an
CC   oral vaccine for botulism) or for the oral delivery of other therapeutic
CC   agents to the general circulation
XX
SQ   Sequence 1291 AA;

  Query Match             93.5%;  Score 2317;  DB 2;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 6.4e-157;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 5
AAY78982
ID   AAY78982 standard; protein; 399 AA.
XX
AC   AAY78982;
XX
DT   20-JUN-2000  (first entry)
XX
DE   C. botulinum type D toxin amino acid sequence.
XX
KW   Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
KW   protect.
XX
OS   Clostridium botulinum.
XX
PN   WO200005252-A1.
XX
PD   03-FEB-2000.
XX
PF   20-JUL-1999;   99WO-IB001301.
XX
PR   22-JUL-1998;   98ZA-00006538.
XX
PA   (AGRI-) AGRIC RES COUNCIL.
XX
PI   De Bruyn EE,  Botha AD;
XX
DR   WPI; 2000-205375/18.
DR   N-PSDB; AAZ98630, AAZ98631.
XX
PT   Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin type
PT   D, useful in vaccines for protection against botulism, comprises at least
PT   one amino acid mutation not present in the wild type D neurotoxins.
XX
PS   Claim 3, 4; Page 54-57; 66pp; English.
XX
CC   This sequence represents the amino acid sequence of a synthetic non-toxic
CC   immunogenic derivative of Clostridium botulinum type D toxin (BoNT).
CC   Botulinum neurotoxin causes botulism poisoning in cattle and sheep, and
CC   usually results in the death of the affected or poisoned animal. The non-
CC   toxic immunogenic fragments of the C. botulinum neurotoxin are useful in
CC   vaccines to protect animals (e.g. humans, cattle, sheep, pigs) against
CC   BoNT type D poisoning. The non-toxic fragments can be produced relatively
CC   simply and inexpensively (specifically by fermentation techniques). As
CC   the fragments are not toxic the risk to production staff is reduced
XX
SQ   Sequence 399 AA;

  Query Match             64.1%;  Score 1589;  DB 3;  Length 399;
  Best Local Similarity   76.0%;  Pred. No. 1.8e-105;
  Matches  307;  Conservative   33;  Mismatches   56;  Indels    8;  Gaps    4;

Qy         62 AEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWV 121
              |||  ||:|||||||||||||||:||||:||||||||||:||||||||||||||||||
Db          1 AEVRVEGNVQLNPIFPFDFKLGSSGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWV 60

Qy        122 SNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFV 181
              ||||||||||||||||||||||||||||||||||||||||:||| ||||||| || |||||||
Db         61 SNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFV 120

Qy        182 TVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWI 241
              |:| |||||| ||||||||||||||||||||||||||::||||:|||||||||||||||
Db        121 TITTNMMGNMMIYINGKLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWI 180

Qy        242 RDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQ 301
              |||||||||| ||||||||||||||||||||||||:|||||:|::|:||||
Db        181 RDFYIFAKELDDKDINILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNG 240

Qy        302 IVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA 361
              ||||||:|||||||||||||||||||||||||| ::|||: ||:|| |:|      ||
Db        241 IVFNTRKNNNDFNEGYKIIIKRIRGNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYK 295

Qy        362 DNHST-EDIYAIGLREQTKD-INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIG 419
              : :   |:  :|  :|  | |     | ||| |   |||||:| |:    |  || |||
Db        296 PSRNLGTDLVPLGALDQPMDEIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIG 355

Qy        420 TYRFRLGGD-WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              :| |:|| | |: | ||:| :|  :||||||||||| ||| ||
Db        356 SYSFKLGDDYWFNHEYLIPVIKIEHYASLLESTSTHWVFVPASE 399


RESULT 6
AAW68398
ID   AAW68398 standard; protein; 451 AA.
XX
AC   AAW68398;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type D toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW   botulism; BotD.
XX
OS   Clostridium botulinum; serotype D.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
```

```
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30591.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 47; Page 352-353; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC   (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC   BotD soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             38.9%;  Score 963.5;  DB 2;  Length 451;
  Best Local Similarity   45.0%;  Pred. No. 1.1e-60;
  Matches  209;  Conservative   82;  Mismatches  146;  Indels   27;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              |||||||||||||||||||||||||||||||||||:||||||||||:|| |||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||    :|||| |:  |||| |||:   |:||  | ||:|:::||: |:|||||:|:|
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKD 117

Qy        121 VSNLPG-YTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKW 178
              ::|    ||||:|:: |||| : | :  : :.|:      :|: | |  | :  || |||
Db        118 LTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKW 177

Qy        179 FFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN 238
              ||||:|||:|| ||:||||:|  : |:::|  :   ||| | |::  |        :|
Db        178 FFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENID--------ENQM 229

Qy        239 MWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYAN 298
              :||||| ||:|||  :||||::      ||:|||||| |::: |||::| :|::||:
Db        230 LWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPE 289

Qy        299 SRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNET 358
              |  :|     : :    | || :     :|:  || :   |  |:: | :    :|
Db        290 SNLLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDT 349

Qy        359 MYADNH---STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGI 415
              :||      |   :||: |:    :     || |: : :   | |||| |:|  ||   :
Db        350 IYATQGGECSQNCVYALKLQSNLGNYGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLL 406

Qy        416 CSI-GTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFV 458
                  |   :||     ::: |    |   || : | |||: | |:

Db        407 ADIYKPWRFS-----FKNAYTPVAVT--NYETKLLSTSSFWKFI 443


RESULT 7
AAB04093
ID   AAB04093 standard; protein; 451 AA.
XX
AC   AAB04093;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype D).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54487.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 6b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 451 AA;

  Query Match             32.3%;  Score 801.5;  DB 4;  Length 451;
  Best Local Similarity   41.7%;  Pred. No. 4.3e-49;
  Matches  183;  Conservative   84;  Mismatches  145;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :|||||||||||:|||||||||||:|| ||||||||||     :|||| |:  |||| ||
Db         26 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
```

```
Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
Db         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
Db        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
Db        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
Db        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKS 314

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHSTEDIYAIGLREQTKD 380
Db        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGDCSQNCVYALKLQSNLGN 374

Qy        381 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 439
Db        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426

Qy        440 KQGNYASLLESTSTHWGFV 458
Db        427 T--NYETKLLSTSSFWKFI 443
```

RESULT 8
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30596.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC express BotG soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX
SQ Sequence 473 AA;

Query Match 29.3%; Score 725.5; DB 2; Length 473;
Best Local Similarity 35.2%; Pred. No. 1.2e-43;
Matches 177; Conservative 82; Mismatches 169; Indels 75; Gaps 17;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA--LLKDIINEYFNNINDSKILSLQNRKNTLVDTS 58
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         59 GYNAEVSEEGDVQLNPIFPPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRIN 118
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNS--ENSNITAHQSKFVVYDSMFDNFSINFWVRTP 118

Qy        119 KWVSN------LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNA 172
Db        119 KYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNI 178

Qy        173 PGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLIT 231
Db        179 SDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKF- 237

Qy        232 SDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYL 291
Db        238 -------VWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQ 290

Qy        292 NRYMYANSRQIVFNTRRNNNDFNE----------GYKIIIKRI---RGNTNDTRVRGGDI 338
Db        291 NIYIKYFSKASMGETAPRTN-FNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDY 349

Qy        339 LYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNN--TY 396
Db        350 IYLNI-----------------DNISDES-YRVYVVLVNSKEIQTQLF--LAPINDDPTF 388

Qy        397 YYASQI----FKSNFNGENIS-------GICSIGTYRFRLGGDWYRH-NYLVPTVKQGNY 444
Db        389 YDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYF--CISQWYL 446

Qy        445 ASLLESTS-----THWGFVPVSE 462
Db        447 RRISENINKLRLGCNWQFIPVDE 469
```

RESULT 9
AAR95009
ID AAR95009 standard; protein; 462 AA.
XX
AC AAR95009;
XX
DT 07-JUL-1996 (first entry)
XX
DE Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /label= Polyhistidine_tag
FT Protein 22. .462

```
FT                   /label= C-fragment
XX
PN   WO9612802-A1.
XX
PD   02-MAY-1996.
XX
PF   23-OCT-1995;   95WO-US013737.
XX
PR   24-OCT-1994;   94US-00329154.
PR   16-MAR-1995;   95US-00405496.
PR   14-APR-1995;   95US-00422711.
PR   07-JUN-1995;   95US-00480604.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Padhye NV,  Kink JA,  Thalley BS,  Stafford DC;
PI   Firca JR;
XX
DR   WPI; 1996-230603/23.
DR   N-PSDB; AAT29246.
XX
PT   Fusion proteins comprising non-toxin protein and part of toxin - useful
PT   to form anti-toxins against Clostridium botulinum type A, and C.
PT   difficile type toxins, and to treat C. difficile intoxication, partic.
PT   diarrhoea.
XX
PS   Claim 7; Page 340-342; 434pp; English.
XX
CC   pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC   (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC   fragment C (see also AAR95008) of the Clostridium botulinum type A
CC   neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC   soluble protein and was purified by metal chelate affinity chromatography
CC   to obtain a product free of endotoxin contamination that may be useful as
CC   an immunogen in vaccine compsns
XX
SQ   Sequence 462 AA;

  Query Match             29.0%;  Score 718.5;  DB 2;  Length 462;
  Best Local Similarity   35.0%;  Pred. No. 3.8e-43;
  Matches  170;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||| |     ||  || :: ||:|:   | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRI 117
               ::::    |  :||       | | ||     |: |     ||||||||:|| ||||||
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS-----KIEVILKNAIVYNSMYENFSTSFWIRI 115

Qy        118 NKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPG 174
               |: :::     ||||: ::||||| : :    :::||:  :: :| : | |    |
Db        116 PKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        175 Y-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSD 233
               | |:| |||:||| : | |||||||:|||   :  |  |: |  | |:::   ||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDT------ 229

Qy        234 SDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR 293
              :  :||: | :| |||: |:|  |:::   : ::||:||: |:|:| |||:|:   |:
Db        230 --HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNK 287

Qy        294 YMYANS-----------RQIVFNTR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILY 340
              |:  |:             |  |  |    |:    | | |||:      |  ||  | :|
Db        288 YVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY 347

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYAS 400
              :: : || | |       | | |   :  |    :  |: :     :    |     :
Db        348 INVVVKNKEYRL--------ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITN 399

Qy        401 QIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTS----THWG 456
              :  | |    | : |  || ::|          | :   |     |   :| :|      |
Db        400 KC-KMNLQDNNGNDIGFIGFHQF---------NNIAKLVASNWYNRQIERSSRTLGCSWE 449

Qy        457 FVPVSE 462
                |:|| :
Db        450 FIPVDD 455


RESULT 10
AAW68390
ID   AAW68390 standard; protein; 462 AA.
XX
AC   AAW68390;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW   Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW   botulism.
XX
OS   Clostridium botulinum; serotype A.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1. .21
FT                    /note= "N-terminal histidine tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30572, AAV30575.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 24; Page 265-267; 428pp; English.
XX
CC   This is the amino acid sequence of a histidine-tagged fragment C
CC   polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC   sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC   AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC   non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on a poly-histidine binding
CC   affinity resin. The invention relates to recombinant proteins derived
CC   from C. botulinum toxins. Methods are provided which allow the isolation
CC   of soluble recombinant proteins that are free of significant endotoxin
CC   contamination. Preferred hosts for production of recombinant proteins are
CC   E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC   immunogens for the production of vaccines and antitoxins that are useful
CC   in the treatment of humans and animals at risk of intoxication with
CC   clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             29.0%;  Score 718.5;  DB 2;  Length 462;
  Best Local Similarity   35.0%;  Pred. No. 3.8e-43;
  Matches  170;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||| |     ||  || :: ||:|:   | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRI 117
```

```
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS-----KIEVILKNAIVYNSMYENFSTSFWIRI 115

Qy        118 NKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPG 174

Db        116 PKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        175 Y-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSD 233

Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDT------ 229

Qy        234 SDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR 293

Db        230 --HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNK 287

Qy        294 YMYANS------------RQIVFNTR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILY 340

Db        288 YVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY 347

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYAS 400

Db        348 INVVVKNKEYRL--------ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITN 399

Qy        401 QIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTS----THWG 456

Db        400 KC-KMNLQDNNGNDIGFIGFHQF---------NNIAKLVASNWYNRQIERSSRTLGCSWE 449

Qy        457 FVPVSE 462

Db        450 FIPVDD 455
```

RESULT 11
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .21
FT   /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997;   97WO-US015394.
XX
PR 28-AUG-1996;   96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 28.8%; Score 715; DB 2; Length 472;
Best Local Similarity 35.8%; Pred. No. 6.9e-43;
Matches 182; Conservative 85; Mismatches 155; Indels 86; Gaps 21;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA--LLKDIINEYFNNINDSKILSLQNRKNTLVDTS 58

Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         59 GYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRIN 118

Db         61 GYGAKVEVYDGVKLND--KNQFKLTSSAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIP 116

Qy        119 KW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNA 172

Db        117 KYRNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176

Qy        173 PGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLIT 231

Db        177 SEYINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLD--------- 226

Qy        232 SDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDY 290

Db        227 GDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGN 286

Qy        291 LNRYMY----ANSRQIVFNTRRNNN-------DFNEGYKIIIKRIRG--NTNDTRVRGGD 337

Db        287 KNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKED 346

Qy        338 ILYFDMTINNK-----AYNLFMKNE-----TMYADNHSTEDIYAIGLREQTKDINDNIIF 387

Db        347 YIHLDLVLHHEEWRVYAYKYFKEQEEKLFLSIISD--SNEFYKTIEIKE----------Y 394

Qy        388 QIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG-------------DWYRHNY 434

Db        395 DEQPS----YSCQLLFKK--DEESTDDIGLIGIHRFYESGVLRKKYKDYFCISKWY---- 444

Qy        435 LVPTVKQGNYASLLESTSTHWGFVPVSE 462

Db        445 -LKEVKRKPYKSNL---GCNWQFIPKDE 468
```

RESULT 12
AAW68396
ID AAW68396 standard; protein; 452 AA.
XX
AC AAW68396;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain NCTC 11219.
OS Synthetic.
XX
FH Key   Location/Qualifiers

```
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;

  Query Match            28.7%;  Score 711.5;  DB 2;  Length 452;
  Best Local Similarity  35.7%;  Pred. No. 1.2e-42;
  Matches  174;  Conservative   84;  Mismatches  156;  Indels   73;  Gaps   18;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAL--------LKDIINEYFNNINDSKILSLQNRKN 52
              ||||||||||||||||||||||||||        |    |::|  |   | :|::: : :
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 60

Qy         53 TLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSIS 112
              ||||||:: ::  |||   |     |  |   :  :| ::||: |:|:: |::||||
Db         61 KYVDTSGYDSNININGDVYKYPTNKNQF--GIDNDKLSEVNISQNDYIIYDNKYKNFSIS 118

Qy        113 FWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDSEQSINFSYD 167
              ||:||    || |:    ||||: :: |||||  : :  | ::::||: |    | :.|:|
Db        119 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYG 178

Qy        168 ISNNAPGY-NKWFFVTVTNNMMGNMKIYIYNGKLIDTIKVKELTGINFSKTITFEINKIPD 226
               :|    | ||| |||:||: :|: |:|||| |||   :  |  |: |  | |:|
Db        179 NANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSY 238

Qy        227 TGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMV 286
              |  |          || | || ||||  :|  |:::   ||::||:||| | |:||||::
Db        239 TRYI---------GIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLL 289

Qy        287 NIDYLNRYM--YANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RGNTNDTRVRGGDILY 340
              |:   | ::    :|   : | |       |    | |: |:|:   :|||  ||:: |:|  .
Db        290 NVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVY 349

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDIND--NIIFQIQPMNNTYYY 398
               :   :   :||    :|||  :|       :|:|  |:   |   |:  ||:
Db        350 INFVASKT--HLF----PLYADTATTN-------KEKTIKISSSGNRFNQVVVMNSVGNN 396

Qy        399 ASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYASLLESTSTH 454
```

```
              :  ||:| || |      ||   |:    :  ||             | : : |:::
Db        397 CTMNFKNN-NGNN------IGLLGFKADTVVASTWY-------------YTHMRDHTNSN 436

Qy        455 ---WGFV 458
                | |:
Db        437 GCFWNFI 443


RESULT 13
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match            28.6%;  Score 709;  DB 2;  Length 451;
  Best Local Similarity  35.1%;  Pred. No. 1.8e-42;
  Matches  171;  Conservative   85;  Mismatches  157;  Indels   74;  Gaps   18;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAL--------LKDIINEYFNNINDSKILSLQNRKN 52
              ||||||||||||||||||||||||||        |    |::|  |   | :|::: : :
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 60
```

```
Qy          53 TLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSIS 112
               ||||||:: ::  |||   |     |  |   :   :| ::||: |:|:: |::||||
Db          61 KYVDTSGYDSNININGDVYKYPTNKNQF--GIYNDKLSEVNISQNDYIIYDNKYKNFSIS 118

Qy         113 FWIRI----NKWVSNLPGYTIIDSVK-NNSGWSIGIISNFLVFTLKQNEDSEQSINFSYD 167
               ||:||    || |:    ||||: :: ||||| : :  | :::||: |    | : |:|
Db         119 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYG 178

Qy         168 ISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPD 226
                :|    | ||| |||:||: :|: |:|||| |||   :  |  |: |  | |:|
Db         179 NANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCSY 238

Qy         227 TGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMV 286
               |  |          || | || ||||  :|  |:|:    |::||:||| | |:||||::
Db         239 TRYI---------GIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLL 289

Qy         287 NI----DYLNRYMYANSRQIVFNTRRN---NNDFNEGYKIIIKRI-RGNTNDTRVRGGDI 338
               |:    :::||    :|   : | |       |    | |: |:|:   :|||  ||  |
Db         290 NVLKPNNFINR--RTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQ 347

Qy         339 LYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYY 398
               :| :  : :| : |       :|||  :|     | :       |  ::       |:
Db         348 VYINF-VASKTHLL-----PLYADTATTNKEKTIKISSSGNRFNQVVVM------NSVGN 395

Qy         399 ASQIFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYASLLESTSTH 454
               :  ||:| || |      ||   |:    :  ||             |  : ::|:::
Db         396 CTMNFKNN-NGNN------IGLLGFKADTVVASTWY------------YTHMRDNTNSN 435

Qy         455 ---WGFV 458
                  | |:
Db         436 GFFWNFI 442
```

RESULT 14
AAW68394
ID AAW68394 standard; protein; 472 AA.
XX
AC AAW68394;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30581.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 303-305; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 28.2%; Score 700; DB 2; Length 472;
Best Local Similarity 36.0%; Pred. No. 8.2e-42;
Matches 182; Conservative 77; Mismatches 166; Indels 80; Gaps 22;

```
Qy           1 MGHHHHHHHHHHSSGHIEGRHMASMA--LLKDIINEYFNNINDSKILSLQNRKNTLVDTS 58
               |||||||||||||||||||||||||  :| :: |:| : | :: ||:|: : | |:| |
Db           1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy          59 GYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRIN 118
               || |:|    |:||      ||| ||    |: ||||:||::||::  ||:||||||
Db          61 GYGAKVEVYDGVELND--KNQFKLTSSA--NSKIRVTQNQNIIFNSVFLDFSVSFWIRIP 116

Qy         119 KW----VSNL--PGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNA 172
               |:    : |     ||||: :|||||| | |  | :::||       :|: | |:|  :
Db         117 KYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDI 176

Qy         173 PGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLIT 231
                 | |:|||||:||| : | ||||||||     :|::  :  :  | |:::
Db         177 SEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLD--------- 226

Qy         232 SDSDNIN-MWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDY 290
                | |    :|:: | ||  ||   :|   :    |:  :||:||| | |||||| |
Db         227 GDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGN 286

Qy         291 LNRYMYANS----RQIVFNTRRNNN-------DFNEGYKIIIKRIRGNT---NDTRVRGG 336
                | |:          :|:  :: | |        |   | | ||:| : |:   ||  ||
Db         287 KNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRR-KSNSQSINDDIVRKE 345

Qy         337 DILYFD-MTINNK----AYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQP 391
               | :| |   :| :     |  | | |          | ::   : :  :  |     ||:
Db         346 DYIYLDFFNLNQEWRVYTYKYFKKEE---------EKLFLAPISDSDEFYN---TIQIKE 393

Qy         392 MNNTYYYASQ-IFKSNFNGENISGICSIGTYRFRLGG------------DWYRHNYLVP 437
                :   |: | :||   : |:   |  || :||   |                ||    :
Db         394 YDEQPTYSCQLLFKK--DEESTDEIGLIGIHRFYESGIVFEEYKDYFCISKWY-----LK 446

Qy         438 TVKQGNYASLLESTSTHWGFVPVSE 462
                ||:  |   |     :| |:|  |
Db         447 EVKRKPYNLKL---GCNWQFIPKDE 468
```

RESULT 15
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX

```
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 48; Page 364-365; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC   sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC   express BotC soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 448 AA;

  Query Match             27.1%;  Score 671.5;  DB 2;  Length 448;
  Best Local Similarity   34.7%;  Pred. No. 8.4e-40;
  Matches  166;  Conservative   80;  Mismatches  177;  Indels   55;  Gaps   15;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||:|    |  :  | || || ::   |  :| |||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
               : :|  |:| :       |  |       :| : || :|:||| |::||||||:|| |
Db         61 GSNISINGNVYIYSTNRNQF--GIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH 118

Qy        121 ---VSNLPGYTIIDSV-KNNSGWSIG---IISNFLVFTLKQNEDSEQSINFSYDISNNAP 173
                 :::   ||||: :  ||||| |    :    :::||:   ::::: | |: |
Db        119 YKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRIS 178

Qy        174 GY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITS 232
              | ||| |||:||| :|| :|||| ||    :  |  |: | | |:         |
Db        179 NYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK---------IVG 229

Qy        233 DSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLN 292
              |   : || | :|  |||  :|  |:::    :::|:|||| | |||:||: |:   :
Db        230 CDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKD 289

Qy        293 RYMYANSRQIVFNTRRN--------NNDFNEGYKIIIKR---IRGNTNDTRVRGGDILYF 341
              :|:  ||  :  | :|          |    || ::||::   |  :  |  ||  |: |
Db        290 KYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYI 349

Qy        342 DMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQ 401
              ::      | |       |||  | ::        :| ::||: : ||  |::     :
Db        350 NVVDRGVEYRL-------YADTKSEKEKII-----RTSNLNDS-LGQIIVMDSIGNNCTM 396

Qy        402 IFKSNFNGENISGICSIGTYRFR----LGGDWYRHNYLVPTVKQGNYASLLESTSTHW 455
              |::| || |      ||   |     :  || :|    |   | : | : |    |
Db        397 NFQNN-NGSN------IGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSI-SKENGW 446


Search completed: September 10, 2004, 03:33:39
Job time : 31.4751 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 9.80605 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 2479
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	718.5	29.0	462	1	US-08-480-604A-26	Sequence 26, Appl
2	718.5	29.0	462	2	US-08-405-496A-26	Sequence 26, Appl
3	718.5	29.0	462	3	US-08-915-136-26	Sequence 26, Appl
4	718.5	29.0	462	4	US-09-084-517-26	Sequence 26, Appl
5	564.5	22.8	438	1	US-08-480-604A-23	Sequence 23, Appl
6	564.5	22.8	438	2	US-08-405-496A-23	Sequence 23, Appl
7	564.5	22.8	438	3	US-08-915-136-23	Sequence 23, Appl
8	564.5	22.8	438	4	US-09-084-517-23	Sequence 23, Appl
9	557.5	22.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
10	557.5	22.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
11	557.5	22.5	1296	3	US-08-915-136-28	Sequence 28, Appl
12	557.5	22.5	1296	4	US-09-084-517-28	Sequence 28, Appl
13	506.5	20.4	1169	4	US-09-255-829-20	Sequence 20, Appl
14	417	16.8	452	1	US-07-618-312A-2	Sequence 2, Appli
15	417	16.8	452	1	US-07-618-312A-4	Sequence 4, Appli
16	417	16.8	452	1	US-08-280-228-2	Sequence 2, Appli
17	417	16.8	452	1	US-08-280-228-4	Sequence 4, Appli
18	417	16.8	618	1	US-08-668-381A-5	Sequence 5, Appli
19	417	16.8	853	4	US-08-913-880C-17	Sequence 17, Appl
20	417	16.8	858	4	US-08-913-880C-16	Sequence 16, Appl
21	417	16.8	860	4	US-08-913-880C-15	Sequence 15, Appl
22	417	16.8	862	4	US-08-913-880C-14	Sequence 14, Appl
23	417	16.8	865	4	US-08-913-880C-13	Sequence 13, Appl
24	417	16.8	866	4	US-08-913-880C-12	Sequence 12, Appl
25	417	16.8	874	4	US-08-913-880C-11	Sequence 11, Appl
26	417	16.8	875	4	US-08-913-880C-10	Sequence 10, Appl
27	417	16.8	1315	4	US-08-913-880C-1	Sequence 1, Appli
28	394	15.9	452	1	US-08-110-786A-8	Sequence 8, Appli
29	153.5	6.2	803	4	US-09-543-681A-4886	Sequence 4886, Ap
30	150	6.1	23	1	US-08-480-604A-24	Sequence 24, Appl
31	150	6.1	23	2	US-08-405-496A-24	Sequence 24, Appl
32	150	6.1	23	3	US-08-915-136-24	Sequence 24, Appl
33	150	6.1	23	4	US-09-084-517-24	Sequence 24, Appl
34	146.5	5.9	977	4	US-09-206-942-53	Sequence 53, Appl
35	146.5	5.9	983	4	US-09-206-942-51	Sequence 51, Appl
36	146	5.9	22	4	US-09-660-742-3	Sequence 3, Appli
37	146	5.9	53	2	US-08-651-818A-19	Sequence 19, Appl
38	146	5.9	53	3	US-09-184-826-19	Sequence 19, Appl
39	146	5.9	54	2	US-08-651-818A-23	Sequence 23, Appl
40	146	5.9	54	3	US-09-184-826-23	Sequence 23, Appl
41	146	5.9	323	1	US-08-591-989-5	Sequence 5, Appli
42	145.5	5.9	476	3	US-09-316-083-3	Sequence 3, Appli
43	145.5	5.9	476	4	US-09-933-700-3	Sequence 3, Appli
44	142.5	5.7	599	2	US-08-910-551B-2	Sequence 2, Appli
45	141	5.7	21	2	US-08-651-818A-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match           29.0%;  Score 718.5;  DB 1;  Length 462;
  Best Local Similarity  35.0%;  Pred. No. 6.9e-51;
  Matches  170;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||| |      ||  || :: ||:|:    | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRI 117
               ::::    |  :||       | | ||     |: |     ||||||||:|| ||||||
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS-----KIEVILKNAIVYNSMYENFSTSFWIRI 115

Qy        118 NKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPG 174
              |: :::     ||||: ::||||| : :    :::||:  :: :| : | |    |
Db        116 PKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        175 Y-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSD 233
              | |:| |||:||| : | ||||||:|||   :  |  |: |  | |:::   ||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDT------ 229

Qy        234 SDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR 293
              :  :||: | :| |||: |:|  |:::   : ::||:||: |:|:| |||:|:   |:
Db        230 --HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNK 287

Qy        294 YMYANS-----------RQIVFNTR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILY 340
              |:  |:           |  |  |    |:    | | |||:       |  ||  | :|
Db        288 YVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY 347

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYAS 400
              :: : || | |           | | |   :  |    :  |: :    :    |     :
Db        348 INVVVKNKEYRL--------ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITN 399

Qy        401 QIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTS----THWG 456
              :  | |    | : |  || ::|         | :   |    |   :| :|      |
Db        400 KC-KMNLQDNNGNDIGFIGFHQF---------NNIAKLVASNWYNRQIERSSRTLGCSWE 449

Qy        457 FVPVSE 462
              |:|| :
Db        450 FIPVDD 455


RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match           29.0%;  Score 718.5;  DB 2;  Length 462;
  Best Local Similarity  35.0%;  Pred. No. 6.9e-51;
  Matches  170;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||| |      ||  || :: ||:|:    | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRI 117
               ::::    |  :||       | | ||     |: |     ||||||||:|| ||||||
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS-----KIEVILKNAIVYNSMYENFSTSFWIRI 115

Qy        118 NKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPG 174
              |: :::     ||||: ::||||| : :    :::||:  :: :| : | |    |
Db        116 PKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        175 Y-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSD 233
              | |:| |||:||| : | ||||||:|||   :  |  |: |  | |:::   ||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDT------ 229

Qy        234 SDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR 293
              :  :||: | :| |||: |:|  |:::   : ::||:||: |:|:| |||:|:   |:
Db        230 --HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNK 287

Qy        294 YMYANS-----------RQIVFNTR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILY 340
              |:  |:           |  |  |    |:    | | |||:       |  ||  | :|
Db        288 YVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY 347

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYAS 400
              :: : || | |           | | |   :  |    :  |: :    :    |     :
Db        348 INVVVKNKEYRL--------ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITN 399

Qy        401 QIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTS----THWG 456
              :  | |    | : |  || ::|         | :   |    |   :| :|      |
Db        400 KC-KMNLQDNNGNDIGFIGFHQF---------NNIAKLVASNWYNRQIERSSRTLGCSWE 449

Qy        457 FVPVSE 462
              |:|| :
Db        450 FIPVDD 455
```

```
RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             29.0%;  Score 718.5;  DB 3;  Length 462;
  Best Local Similarity   35.0%;  Pred. No. 6.9e-51;
  Matches  170;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||| |      ||  || :: ||:|:   | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRI 117
                ::::    |  :||       | | ||     |: |     ||||||||:|| ||||||
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS-----KIEVILKNAIVYNSMYENFSTSFWIRI 115

Qy        118 NKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPG 174
              |: :::     ||||: ::||||| : :    :::||:  :: :| : | |     |
Db        116 PKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        175 Y-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSD 233
              | |:| |||:||| : | ||||||:|||   :  | :|: |  | |:::   ||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDT------ 229

Qy        234 SDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR 293
              :  :||: | :| |||: |:|  |:::   : ::||:||: |:|:| |||:|:   |:
Db        230 --HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNK 287

Qy        294 YMYANS-----------RQIVFNTR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILY 340
              |:  |:             |  |  |    |:    | | |||:      |  ||  | :|
Db        288 YVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY 347

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYAS 400
              :: : || | |           | | |   :  |   :  |: :    :    |    :
Db        348 INVVVKNKEYRL--------ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITN 399

Qy        401 QIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTS----THWG 456
              :  | |    | : |  || ::|         | :   |    |   :| :|      |
Db        400 KC-KMNLQDNNGNDIGFIGFHQF---------NNIAKLVASNWYNRQIERSSRTLGCSWE 449

Qy        457 FVPVSE 462
              |:|| :
Db        450 FIPVDD 455


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
```

```
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             29.0%;  Score 718.5;  DB 4;  Length 462;
  Best Local Similarity   35.0%;  Pred. No. 6.9e-51;
  Matches  170;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              |||||||||||||||||||||||||| |      ||  || :: ||:|:    | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVSEEGDVQLNPIFPFD---FKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRI 117
               ::::    |  :||       | | ||      |: |     ||||||||:|| ||||||
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS-----KIEVILKNAIVYNSMYENFSTSFWIRI 115

Qy        118 NKWVSNLP---GYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPG 174
               |: :::     ||||: ::||||| : :    :::||:  :: :| : | |      |
Db        116 PKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        175 Y-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSD 233
              | |:| |||:||| : | ||||||:|||   :  |  |: |  | |:::   ||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDT------ 229

Qy        234 SDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNR 293
              :  :||: | :| ||| |:|  |::   : ::||:||: |:|:| |||:|:   |:
Db        230 --HRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNK 287

Qy        294 YMYANS------------RQIVFNTR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILY 340
              |:  |:            |  | |   |:   | | |||:      |  ||  | :|
Db        288 YVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY 347

Qy        341 FDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYAS 400
               :: : || | |         | | |   :  |    :  |: :     :    |     :
Db        348 INVVVKNKEYRL--------ATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITN 399

Qy        401 QIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTS----THWG 456
              :  | |    | : |  || ::|          | :   |     :| :|       |
Db        400 KC-KMNLQDNNGNDIGFIGFHQF---------NNIAKLVASNWYNRQIERSSRTLGCSWE 449

Qy        457 FVPVSE 462
              |:|| :
Db        450 FIPVDD 455


RESULT 5
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match             22.8%;  Score 564.5;  DB 1;  Length 438;
  Best Local Similarity   31.6%;  Pred. No. 2.7e-38;
  Matches  146;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy         25 MALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FK 81
              || |      ||  || :: ||:|:    | |:| | | ::::    |  :||       |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSG 138
              | ||      |: |     |||||||:|| |||||| |: :::     ||||: ::||||
Db         61 LESS-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSG 115

Qy        139 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 197
              | : :    :::||:  :: :| : | |      |   | |:| |||:||| : | ||||||
Db        116 WKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYING 175

Qy        198 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 257
              :|||   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|
Db        176 RLIDQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIK 227

Qy        258 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFN 305
              |:::   : ::||:||: |:|:| |||:|:   |:|:  |:             | |
Db        228 DLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMT 287

Qy        306 TR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 364
              |    |:    | | |||:      |  ||  | :| :: : || | |        | | |
Db        288 TNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNA 339

Qy        365 STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR 424
              |   :  |    :  |: :     :     |     ::  | |     | : |  || ::|
```

```
Db        340 SQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF- 397

Qy        425 LGGDWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                  | :   |    |   :| :|      | |:|| :
Db        398 --------NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 431


RESULT 6
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match             22.8%;  Score 564.5;  DB 2;  Length 438;
  Best Local Similarity   31.6%;  Pred. No. 2.7e-38;
  Matches  146;  Conservative   80;  Mismatches  181;  Indels   55;  Gaps   11;

Qy         25 MALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FK 81
                 || |      ||  || :: ||:|:    | |:| | | ::::    |  :||       |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSG 138
                | ||     |: |     ||||||||:|| |||||| |: :::     ||||: ::||||
Db         61 LESS-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSG 115

Qy        139 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 197
                | : :    :::||:  :: :| : | |    |   | |:| |||:||| : | ||||||

Db        116 WKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYING 175

Qy        198 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 257
                :|||   :  |  |: |  | |:::   ||       :  :||: | :| |||: |:|
Db        176 RLIDQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIK 227

Qy        258 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS-----------RQIVFN 305
                |:::   : ::||:||: |:|:| |||:|:   |:|:  |:            |  |
Db        228 DLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMT 287

Qy        306 TR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 364
                |    |:    | | |||:      |  ||  | :| :: : || | |         | |
Db        288 TNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNA 339

Qy        365 STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR 424
                |  : |    :  |: :     :     |     ::  | |    | : |  || ::|
Db        340 SQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF- 397

Qy        425 LGGDWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                  | :   |    |   :| :|      | |:|| :
Db        398 --------NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 431


RESULT 7
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
```

; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-23

Query Match 22.8%; Score 564.5; DB 3; Length 438;
Best Local Similarity 31.6%; Pred. No. 2.7e-38;
Matches 146; Conservative 80; Mismatches 181; Indels 55; Gaps 11;

```
Qy         25 MALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FK 81
              || |      ||  || :: ||:|:   | |:| | | ::::    | :||       |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSG 138
              | ||      |: |     |||||||:|| |||||| |: :::     ||||: ::||||
Db         61 LESS-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSG 115

Qy        139 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 197
              | : :    :::||:  :: :| : | |     |  | |:| |||:||| : | ||||||
Db        116 WKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYING 175

Qy        198 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 257
              :|||   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|
Db        176 RLIDQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIK 227

Qy        258 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFN 305
              |:::   : ::||:||: |:|:| |||:|:   |:|:  |:             | |
Db        228 DLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMT 287

Qy        306 TR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 364
              |   |:   | | |||:      |  ||  | :| :: : || | |          | |
Db        288 TNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNA 339

Qy        365 STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR 424
              |   :  |   :  |: :     :    |     ::  | |    | : |  || ::|
Db        340 SQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF- 397

Qy        425 LGGDWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                  | :  |    |   :| :|      | |:|| :
Db        398 --------NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 431
```

RESULT 8
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-23

Query Match 22.8%; Score 564.5; DB 4; Length 438;
Best Local Similarity 31.6%; Pred. No. 2.7e-38;
Matches 146; Conservative 80; Mismatches 181; Indels 55; Gaps 11;

```
Qy         25 MALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FK 81
              || |      ||  || :: ||:|:   | |:| | | ::::    | :||       |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSG 138
              | ||      |: |     |||||||:|| |||||| |: :::     ||||: ::||||
Db         61 LESS-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSG 115

Qy        139 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 197
              | : :    :::||:  :: :| : | |     |  | |:| |||:||| : | ||||||
Db        116 WKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYING 175

Qy        198 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 257
              :|||   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|
Db        176 RLIDQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIK 227

Qy        258 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFN 305
              |:::   : ::||:||: |:|:| |||:|:   |:|:  |:             | |
Db        228 DLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMT 287

Qy        306 TR-RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 364
              |   |:   | | |||:      |  ||  | :| :: : || | |          | |
Db        288 TNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNA 339

Qy        365 STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR 424
              |   :  |   :  |: :     :    |     ::  | |    | : |  || ::|
Db        340 SQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF- 397

Qy        425 LGGDWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                  | :  |    |   :| :|      | |:|| :
Db        398 --------NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 431
```

RESULT 9
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139

```
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             22.5%;  Score 557.5;  DB 1;  Length 1296;
  Best Local Similarity   31.4%;  Pred. No. 4.3e-37;
  Matches  144;  Conservative   80;  Mismatches  180;  Indels   55;  Gaps   11;

Qy         28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGS 84
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 141
Db        922 S-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 976

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLY 1088

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR- 307
Db       1089 DNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNI 1148

Qy        308 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 367
Db       1149 YLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQA 1200

Qy        368 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 427
Db       1201 GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---- 1255

Qy        428 DWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
Db       1256 -----NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289


RESULT 10
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28
```

```
  Query Match             22.5%;  Score 557.5;  DB 2;  Length 1296;
  Best Local Similarity   31.4%;  Pred. No. 4.3e-37;
  Matches  144;  Conservative   80;  Mismatches  180;  Indels   55;  Gaps   11;

Qy         28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGS 84
              |     ||  || :: ||:|:   | |:| | | ::::    |  :||       | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 141
              |     |: |     ||||||||:|| |||||| |: :::     ||||: ::||||| :
Db        922 S-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 976

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :   :::||:  :: :| : | |     |   | |:| |||:||| : | |||||:||
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|  |:
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLY 1088

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR- 307
              ::   : ::||:||: |:|:| |||:|:   |:|:  |:              | | |
Db       1089 DNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNI 1148

Qy        308 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 367
              |:    | | |||:       |  ||  | :| :: : || | |        | | |
Db       1149 YLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQA 1200

Qy        368 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 427
              :  |    :  |: :    :    |     ::  | |     | : |  || ::|
Db       1201 GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---- 1255

Qy        428 DWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                 | :   |    |   :| :|      | |:|| :
Db       1256 -----NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289


RESULT 11
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             22.5%;  Score 557.5;  DB 3;  Length 1296;
  Best Local Similarity   31.4%;  Pred. No. 4.3e-37;
  Matches  144;  Conservative   80;  Mismatches  180;  Indels   55;  Gaps   11;

Qy         28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGS 84
              |     ||  || :: ||:|:   | |:| | | ::::    |  :||       | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 141
              |     |: |     ||||||||:|| |||||| |: :::     ||||: ::||||| :
Db        922 S-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 976

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :   :::||:  :: :| : | |     |   | |:| |||:||| : | |||||:||
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|  |:
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLY 1088

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR- 307
              ::   : ::||:||: |:|:| |||:|:   |:|:  |:              | | |
Db       1089 DNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNI 1148

Qy        308 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 367
              |:    | | |||:       |  ||  | :| :: : || | |        | | |
Db       1149 YLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQA 1200

Qy        368 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 427
              :  |    :  |: :    :    |     ::  | |     | : |  || ::|
Db       1201 GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---- 1255

Qy        428 DWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                 | :   |    |   :| :|      | |:|| :
Db       1256 -----NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289


RESULT 12
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
```

```
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match           22.5%;  Score 557.5;  DB 4;  Length 1296;
  Best Local Similarity  31.4%;  Pred. No. 4.3e-37;
  Matches  144;  Conservative   80;  Mismatches  180;  Indels   55;  Gaps   11;

Qy         28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGS 84
              |     ||  || :: ||:|:   | |:| | | ::::    |  :||       | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 141
              |     |: |     ||||||||:|| |||||| |: :::     ||||: ::||||| :
Db        922 S-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 976

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :   :::||:  :: : | : | |    |  | |:| |||:||| : | ||||||:||
Db        977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|  |:
Db       1037 DQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLY 1088

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR- 307
              ::   : ::||:||: |:|:| |||:|:   |:|:  |:             |  |  |
Db       1089 DNQSNSGILKDFWGDYLQYDKPYYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNI 1148

Qy        308 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 367
                |:    | | |||:      |  ||  | :| :: : || | |        | | |
```

```
Db       1149 YLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQA 1200

Qy        368 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 427
              : |   :  |: :   :     |    ::  | |    | : |  || ::|
Db       1201 GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---- 1255

Qy        428 DWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                | :   |     |   :| :|       | |:|| :
Db       1256 -----NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289


RESULT 13
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-20

  Query Match           20.4%;  Score 506.5;  DB 4;  Length 1169;
  Best Local Similarity  37.2%;  Pred. No. 5.6e-33;
  Matches  125;  Conservative   59;  Mismatches  115;  Indels   37;  Gaps   12;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :| :: |:| : | :: ||:|: : | |:| ||| |:|     |:||       ||| ||
Db        848 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA 905

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140
                 |: ||||:||::||::  ||:|||||| |:    : |     ||||: :||||||
Db        906 --NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
              | |  | :::||       :|: | |:|  :   | |:|||||:||| : | ||||||||
Db        964 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 1022
```

```
Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINI 258
              :|::  :  :  | |:::          | |    :|:: | ||  ||   :|
Db       1023 ESNTDIKDIREVIANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIEE 1073

Qy        259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN--- 311
               :    |:  :||:||| | |||||| |     | |:          :|:  :: | |
Db       1074 RYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKY 1133

Qy        312 ----DFNEGYKIIIKRIRGNT---NDTRVRGGDILY 340
                  |  | | ||:| : |:   ||  ||  | :|
Db       1134 INYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIY 1168


RESULT 14
US-07-618-312A-2
; Sequence 2, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             16.8%;  Score 417;  DB 1;  Length 452;
  Best Local Similarity   26.5%;  Pred. No. 3.4e-26;
  Matches  127;  Conservative   86;  Mismatches  162;  Indels  104;  Gaps   19;

Qy         39 INDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGK 91
              :  | ||:|    :  :| ||:|: |    | || |         | :|:       :  :
Db         18 LKKSTILNLDINNDINIDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSE 69

Qy         92 VIVTQNENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWS 140
              ||| :  :| || |: :|::|||:|: |   |:|       :|| |:| :     ||||
Db         70 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 129

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGK 198
              : :  | |::||| :    : | |  |: :    |  ||| |:|:||: : :  :||||
Db        130 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 188

Qy        199 LIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINI 258
              |: : ::  |  |      || ::::         :::|  :  |  | || | |:|
Db        189 LMGSAEITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEK 240

Qy        259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNI------------DYL---NRYMYANSRQI 302
              |: |       ::|:||| |||: |||:: :           ||:   |  | | :
Db        241 LYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLN 300

Qy        303 VFNTRRNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLF 353
              ::  |  |     | | ||||    |   |: |: || :   :: ||  :          | |
Db        301 IYYRRLYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAF 355

Qy        354 MKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 413
               : :    ::  |       | |              : :   |: |:    :: :| |
Db        356 NNLDRILRVGYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL--KLYDDKNAS 401

Qy        414 GICSIGTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
               :  :||:  ::| |           ||  |:|  :                | |||  |
Db        402 -LGLVGTHNGQIGNDPNRDILIASNWY-FNHLKDKI----------LGCDWYFVPTDE 447


RESULT 15
US-07-618-312A-4
; Sequence 4, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-07-618-312A-4

Query Match 16.8%; Score 417; DB 1; Length 452;
Best Local Similarity 26.5%; Pred. No. 3.4e-26;
Matches 127; Conservative 86; Mismatches 162; Indels 104; Gaps 19;

```
Qy         39 INDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGK 91
              :  | ||:|    : : | ||:|: |    | || |         | :|:       :  :
Db         18 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSE 69

Qy         92 VIVTQNENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWS 140
              ||| :  :| || |: :|::|||:|: |   |:|       |:|| |:| :     ||||
Db         70 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 129

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGK 198
              : :  | |::||| :    : | |  |: :    |  ||| |:|:||: : :  :||||
Db        130 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 188

Qy        199 LIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINI 258
              |: : ::  |  |     || ::::        :::|  : |  | || | |: |:|
Db        189 LMGSAEITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEK 240

Qy        259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNI-------------DYL---NRYMYANSRQI 302
              |: |      ::|:||| |||: |||:: :             ||:   |   | | :
Db        241 LYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLN 300

Qy        303 VFNTRRNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLF 353
              ::  |  |     | | ||||   |   |: |: || :   :: ||  :        | |
Db        301 IYYRRLYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAF 355

Qy        354 MKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 413
             : :   ::   |      | |           : :   |: |:    :: :| |
Db        356 NNLDRILRVGYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL--KLYDDKNAS 401

Qy        414 GICSIGTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              :  :||:  ::| |          ||  |:|  :               | |||  |
Db        402 -LGLVGTHNGQIGNDPNRDILIASNWY-FNHLKDKI----------LGCDWYFVPTDE 447
```

Search completed: September 10, 2004, 03:47:09
Job time : 10.806 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 89.4288 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 2479
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2479	100.0	462	12	US-10-272-898-62	Sequence 62, Appl
2	2479	100.0	462	15	US-10-354-774-62	Sequence 62, Appl
3	2479	100.0	462	15	US-10-271-012-62	Sequence 62, Appl
4	2479	100.0	462	16	US-10-729-122-62	Sequence 62, Appl
5	2479	100.0	462	16	US-10-729-039-62	Sequence 62, Appl
6	2317	93.5	450	10	US-09-910-186A-10	Sequence 10, Appl
7	2317	93.5	1291	12	US-10-272-898-60	Sequence 60, Appl
8	2317	93.5	1291	15	US-10-354-774-60	Sequence 60, Appl
9	2317	93.5	1291	15	US-10-271-012-60	Sequence 60, Appl
10	2317	93.5	1291	15	US-10-452-024-3	Sequence 3, Appli
11	2317	93.5	1291	15	US-10-452-024-135	Sequence 135, App
12	2317	93.5	1291	15	US-10-452-024-136	Sequence 136, App
13	2317	93.5	1291	15	US-10-205-516-6	Sequence 6, Appli
14	2317	93.5	1291	16	US-10-729-122-60	Sequence 60, Appl
15	2317	93.5	1291	16	US-10-729-039-60	Sequence 60, Appl
16	2317	93.5	1301	15	US-10-205-516-20	Sequence 20, Appl
17	1755	70.8	1285	15	US-10-452-024-165	Sequence 165, App
18	1754	70.8	1285	15	US-10-452-024-166	Sequence 166, App
19	963.5	38.9	451	12	US-10-272-898-68	Sequence 68, Appl
20	963.5	38.9	451	15	US-10-354-774-68	Sequence 68, Appl
21	963.5	38.9	451	15	US-10-271-012-68	Sequence 68, Appl
22	963.5	38.9	451	16	US-10-729-122-68	Sequence 68, Appl
23	963.5	38.9	451	16	US-10-729-039-68	Sequence 68, Appl
24	813.5	32.8	1280	15	US-10-452-024-139	Sequence 139, App
25	813.5	32.8	1280	15	US-10-452-024-140	Sequence 140, App
26	801.5	32.3	1275	15	US-10-452-024-167	Sequence 167, App
27	801.5	32.3	1276	12	US-10-272-898-66	Sequence 66, Appl
28	801.5	32.3	1276	15	US-10-354-774-66	Sequence 66, Appl
29	801.5	32.3	1276	15	US-10-271-012-66	Sequence 66, Appl
30	801.5	32.3	1276	15	US-10-452-024-4	Sequence 4, Appli
31	801.5	32.3	1276	15	US-10-452-024-164	Sequence 164, App
32	801.5	32.3	1276	15	US-10-452-024-168	Sequence 168, App
33	801.5	32.3	1276	15	US-10-205-516-8	Sequence 8, Appli
34	801.5	32.3	1276	16	US-10-729-122-66	Sequence 66, Appl
35	801.5	32.3	1276	16	US-10-729-039-66	Sequence 66, Appl
36	801.5	32.3	1286	15	US-10-205-516-22	Sequence 22, Appl
37	798.5	32.2	451	10	US-09-910-186A-12	Sequence 12, Appl
38	725.5	29.3	473	12	US-10-272-898-79	Sequence 79, Appl
39	725.5	29.3	473	15	US-10-354-774-79	Sequence 79, Appl
40	725.5	29.3	473	15	US-10-271-012-79	Sequence 79, Appl
41	725.5	29.3	473	16	US-10-729-122-79	Sequence 79, Appl
42	725.5	29.3	473	16	US-10-729-039-79	Sequence 79, Appl
43	718.5	29.0	462	12	US-10-272-898-26	Sequence 26, Appl
44	718.5	29.0	462	15	US-10-354-774-26	Sequence 26, Appl
45	718.5	29.0	462	15	US-10-271-012-26	Sequence 26, Appl

ALIGNMENTS

```
RESULT 1
US-10-272-898-62
; Sequence 62, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
```

; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-272-898-62

Query Match 100.0%; Score 2479; DB 12; Length 462;
Best Local Similarity 100.0%; Pred. No. 2e-193;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
```

RESULT 2
US-10-354-774-62
; Sequence 62, Application US/10354774
; Publication No. US20030215468A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/354,774
; FILING DATE: 30-Jan-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-354-774-62

Query Match 100.0%; Score 2479; DB 15; Length 462;
Best Local Similarity 100.0%; Pred. No. 2e-193;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
```

RESULT 3
US-10-271-012-62
; Sequence 62, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 62:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 462 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-271-012-62

  Query Match           100.0%;  Score 2479;  DB 15;  Length 462;
  Best Local Similarity  100.0%;  Pred. No. 2e-193;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 4
US-10-729-122-62
; Sequence 62, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 62:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 462 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-729-122-62

  Query Match           100.0%;  Score 2479;  DB 16;  Length 462;
  Best Local Similarity  100.0%;  Pred. No. 2e-193;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 5
US-10-729-039-62
```

```
; Sequence 62, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 62:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 462 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-729-039-62

  Query Match           100.0%;  Score 2479;  DB 16;  Length 462;
  Best Local Similarity  100.0%;  Pred. No. 2e-193;
  Matches  462;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFF 180

Qy        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMW 240

Qy        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSR 300

Qy        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMY 360

Qy        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGT 420

Qy        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
              |||||||||||||||||||||||||||||||||||||||||||
Db        421 YRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462


RESULT 6
US-09-910-186A-10
; Sequence 10, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 450
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
;   OTHER INFORMATION: sequence
US-09-910-186A-10

  Query Match           93.5%;  Score 2317;  DB 10;  Length 450;
  Best Local Similarity  99.8%;  Pred. No. 3.1e-180;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 73

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 133

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 193

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 253

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 313

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        314 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 373

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 433

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db        434 SLLESTSTHWGFVPVSE 450


RESULT 7
US-10-272-898-60
; Sequence 60, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-272-898-60

  Query Match             93.5%;  Score 2317;  DB 12;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 8
US-10-354-774-60
; Sequence 60, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-354-774-60

  Query Match             93.5%;  Score 2317;  DB 15;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914
```

```
Qy          86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy         146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy         206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy         266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy         326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy         386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy         446 SLLESTSTHWGFVPVSE 462
               |||||||||||||||||
Db        1275 SLLESTSTHWGFVPVSE 1291
```

RESULT 9
US-10-271-012-60
; Sequence 60, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-271-012-60

```
  Query Match           93.5%;  Score 2317;  DB 15;  Length 1291;
  Best Local Similarity  99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
               :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy          86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy         146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy         206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy         266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy         326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy         386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy         446 SLLESTSTHWGFVPVSE 462
               |||||||||||||||||
Db        1275 SLLESTSTHWGFVPVSE 1291
```

RESULT 10
US-10-452-024-3
; Sequence 3, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-3

```
  Query Match           93.5%;  Score 2317;  DB 15;  Length 1291;
  Best Local Similarity  99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
               :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy          86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy         146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
                 |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 11
US-10-452-024-135
; Sequence 135, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 135
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum phage
US-10-452-024-135

  Query Match             93.5%;  Score 2317;  DB 15;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
                :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
                 |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 12
US-10-452-024-136
; Sequence 136, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 136
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-136

  Query Match             93.5%;  Score 2317;  DB 15;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
                :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
                 |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 13
US-10-205-516-6
; Sequence 6, Application US/10205516
; Publication No. US20040018589A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-6

  Query Match             93.5%;  Score 2317;  DB 15;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 14
US-10-729-122-60
; Sequence 60, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-122-60

  Query Match             93.5%;  Score 2317;  DB 16;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 15
US-10-729-039-60
; Sequence 60, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
```

```
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 60:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1291 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-729-039-60

  Query Match            93.5%;  Score 2317;  DB 16;  Length 1291;
  Best Local Similarity  99.8%;  Pred. No. 1.2e-179;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


Search completed: September 10, 2004, 04:58:15
Job time : 90.4288 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 8.16192 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 2479
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2317	93.5	1291	2	S46431	botulinum neurotox
2	2317	93.5	1291	2	A49777	botulinum neurotox
3	1754	70.8	1285	2	S70582	botulinum neurotox
4	801.5	32.3	1276	2	S11455	botulinum neurotox
5	573.5	23.1	1297	2	S39791	neurotoxin - Clost
6	565	22.8	1291	2	I40631	non-proteolytic bo
7	558.5	22.5	1252	2	S21178	botulinum neurotox
8	558	22.5	1251	2	JH0256	botulinum neurotox
9	557.5	22.5	1296	1	BTCLAB	bontoxilysin (EC 3
10	548	22.1	1291	1	A48940	bontoxilysin (EC 3
11	532	21.5	1268	2	S33411	botulinum neurotox
12	525.5	21.2	1296	2	I40645	botulinum neurotox
13	509.5	20.6	1274	2	I40813	neurotoxin type F
14	417	16.8	1315	1	BTCLTN	tentoxilysin (EC 3
15	245.5	9.9	367	2	S48106	neurotoxin type E
16	240	9.7	366	2	S48110	neurotoxin type F
17	222	9.0	1162	2	A47708	progenitor toxin n
18	222	9.0	1162	2	I40817	botulinum toxin no
19	213	8.6	1193	2	S68218	botulinum neurotox
20	205.5	8.3	369	2	S48109	neurotoxin type F
21	196	7.9	1196	2	JQ1467	toxin, nontoxic co
22	196	7.9	1196	2	S46430	botulinum neurotox
23	195.5	7.9	1165	2	I40644	botulinum neurotox
24	187	7.5	1193	2	JC4901	nontoxic-nonhemagg
25	171.5	6.9	960	2	S72284	DNA-directed RNA p
26	167.5	6.8	398	1	R3BYM1	ribosomal protein
27	167	6.7	4688	2	F82885	hypothetical prote
28	166.5	6.7	1844	2	D71612	hypothetical prote
29	163.5	6.6	617	2	D96978	hypothetical prote
30	162.5	6.6	1817	2	H71611	probable secreted
31	158	6.4	1711	2	T18429	hypothetical prote
32	157.5	6.4	762	2	E81371	probable periplasm
33	154	6.2	2817	2	B97033	uncharacterized pr
34	153.5	6.2	3724	2	T18427	hypothetical prote
35	153	6.2	1306	2	T28313	ORF MSV152 probabl
36	151	6.1	1182	2	G71607	probable integral
37	149.5	6.0	3973	2	B71612	hypothetical prote
38	148	6.0	588	2	A71661	hypothetical prote
39	146.5	5.9	2013	2	C71610	probable membrane
40	146	5.9	1365	2	T18419	hypothetical prote
41	145.5	5.9	870	2	S27514	mosquitocidal toxi
42	145.5	5.9	1465	2	A70199	hypothetical prote
43	145.5	5.9	1943	2	B64596	toxin-like outer m
44	144.5	5.8	2485	1	H71621	serine/threonine-s
45	144	5.8	883	2	T18484	hypothetical prote

ALIGNMENTS

RESULT 1
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 93.5%; Score 2317; DB 2; Length 1291;
Best Local Similarity 99.8%; Pred. No. 7.9e-126;
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274
```

```
Qy        446 SLLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291
```

RESULT 2
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 93.5%; Score 2317; DB 2; Length 1291;
  Best Local Similarity 99.8%; Pred. No. 7.9e-126;
  Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              ||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291
```

RESULT 3
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C anc
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membra
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match 70.8%; Score 1754; DB 2; Length 1285;
  Best Local Similarity 76.8%; Pred. No. 2.1e-93;
  Matches 338; Conservative 38; Mismatches 56; Indels 8; Gaps 4;

```
Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||:||||||:||||||||||:||||:|||||||  ||:||||||||||||||||
Db        851 SLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSS 910

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              |:||||:||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db        911 GDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 970

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||:||| |||||||| || |||||||||:| |||||| ||||||||||||
Db        971 NFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFVTITTNMMGNMMIYINGKLIDTIKV 1030

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||::||||:|||||||||||||||||||||||||||||||||||||
Db       1031 KELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSLQY 1090

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||:|||||:|::|:||||      ||||||:||||||||||||||||
Db       1091 TNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFNEGYKIIIKRIR 1150

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAIGLREQTKD-IND 383
              |||||||||  ::|||: ||:|| |:|      ||  : :   |:  :|  :|  | |
Db       1151 GNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYKPSRNLGTDLVPLGALDQPMDEIRK 1205

Qy        384 NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGD-WYRHNYLVPTVKQG 442
                | ||| |   |||||:| |:    |  || |||:| |:|| | |: | ||:| :|
Db       1206 YGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLGDDYWFNHEYLIPVIKIE 1265

Qy        443 NYASLLESTSTHWGFVPVSE 462
              :||||||||||| ||| ||
Db       1266 HYASLLESTSTHWVFVPASE 1285
```

```
RESULT 4
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             32.3%;  Score 801.5;  DB 2;  Length 1276;
  Best Local Similarity   41.7%;  Pred. No. 1.5e-38;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :|||||||||||:|||||||||||:|| ||||||||||     :|||| |:  |||| ||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        911 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 967

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        968 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 1027

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
              |:::|  :   ||| | |::  | .      :|  :|||| ||:|||  :||||::
Db       1028 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 1079

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
              ||:|||||| |::: |||::| :|::||:   |  :|     : :    |  | ||
Db       1080 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKS 1139

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 380
              :      :|:  || :   |  |:: | :    :|:||      |   :||: |:    :
Db       1140 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 1199

Qy        381 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 439
                 || |: : :   | |||| |:|  ||   :  |   :||      ::: |     |
Db       1200 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 1251

Qy        440 KQGNYASLLESTSTHWGFV 458
                || : | |||: | |:
Db       1252 T--NYETKLLSTSSFWKFI 1268


RESULT 5
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             23.1%;  Score 573.5;  DB 2;  Length 1297;
  Best Local Similarity   31.8%;  Pred. No. 2.1e-25;
  Matches  151;  Conservative   82;  Mismatches  169;  Indels   73;  Gaps   16;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|  : | | :||: : ||||  |   |:|:||| | ::   ||  | |   ||| :|
Db        853 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNS- 911

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSGWS 140
               :   :   |:: :||:||:::|||:||:|  |: :|         ||||  :||:|||
Db        912 -ENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
              : |  | :::||        :|| | | | :|   | |||| :|:||: :||  ||||| |
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINIL 259
              : |:  |  || |  | |::    ||            :||:|| || :||:  ::: |
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKF--------VWIKDFNIFGRELNATEVSSL 1082

Qy        260 FNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE---- 315
              :     || :||:||| |||: :||: |    | |:   |:  : |    | ||
Db       1083 YWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAAIN 1141

Qy        316 ------GYKIIIKRI---RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST 366
                   | : |||:    |   ||  || || :| ::                  || |
Db       1142 YQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI----------------DNISD 1183

Qy        367 EDIYAIGLREQTKDINDNIIFQIQPMNN--TYYYASQI----FKSNFNGENIS------- 413
              |  | : :   :|:|   :   : |:|:  |:|    ||      |: :| : :
Db       1184 ES-YRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTF 1240

Qy        414 GICSIGTYRFRLGGDWYRH-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 462
              |:  || :    |  |  : ||    : |     : |: :       :| |:|| |
Db       1241 GLFGIGKFVKDYGYVWDTYDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 1293


RESULT 6
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
```

C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 22.8%; Score 565; DB 2; Length 1291;
Best Local Similarity 32.4%; Pred. No. 6.6e-25;
Matches 157; Conservative 86; Mismatches 157; Indels 84; Gaps 20;

```
Qy         23 ASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKL 82
              ::: :|  | |:| : | :: ||:|: | | |:| ||| |:|     |:||      |||
Db        844 SNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKL 901

Qy         83 GSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNN 136
               || :   |: ||||:||::|||:  ||:|||||| |:    : |     ||||: :|||
Db        902 TSSAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNN 959

Qy        137 SGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYI 195
              ||| | |  | :::||       :|: | |:|  :   | |:|||||:||| : | ||||
Db        960 SGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYI 1018

Qy        196 NGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGK 254
              || |   : :|::  :  :  |||:::          | |    :|:: | ||  :|:
Db       1019 NGTLESNMDIKDIGEVIVNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQS 1069

Qy        255 DINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNN 310
              :|  ::    |:  :||:||| | ||||||| |    | |:     ::  :|:  :: |
Db       1070 NIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQ 1129

Qy        311 N-------DFNEGYKIIIKRIRG--NTNDTRVRGGDILYFDMTINNK-----AYNLFMKN 356
              |       :  | | ||:|     : ||  ||  | :: |: ::::     ||  | :
Db       1130 NSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQ 1189

Qy        357 E-----TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGEN 411
              |     :: :|  | |     | ::|          :  ||      |    :||   : |:
Db       1190 EEKLFLSIISD--SNEFYKTIEIKE----------YDEQPS----YSCQLLFKK--DEES 1231

Qy        412 ISGICSIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFV 458
                |  || :||   |              ||     :  ||:  | | |      :| |:
Db       1232 TDDIGLIGIHRFYESGVLRKKYKDYFCISKWY-----LKEVKRKPYKSNL---GCNWQFI 1283

Qy        459 PVSE 462
              | |
Db       1284 PKDE 1287
```

RESULT 7
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Prevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other c
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

Query Match 22.5%; Score 558.5; DB 2; Length 1252;
Best Local Similarity 32.5%; Pred. No. 1.5e-24;
Matches 147; Conservative 85; Mismatches 156; Indels 65; Gaps 17;

```
Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    |::|  |  | :|:::  : :  ||||||:: ::  |||    |     | |
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 892

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
              :   :| ::||: |:|::  |::||||||:||    || |:     ||||: :: ||||| :
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 952

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :  | :::||: |     | : |:|  :|     | ||| |||:||: :|: |:|||| ||
Db        953 SLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLI 1012

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:|      |  |            || | || ||||  :|  |:
Db       1013 DQKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1063

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYM--YANSRQIVFNTRRN---NNDFNE 315
              ::   ||::||:||| | |:|||::|:   | ::    :|   : | |       |
Db       1064 SNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYS 1123

Qy        316 GYKIIIKRI-RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGL 374
              | |: |:|:   :|||  ||  | :| :   :    :||     :|||  :|
Db       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKT--HLF----PLYADTATTN------- 1170

Qy        375 REQTKDIND--NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGD 428
              :|:|  |:   |  |:  ||:    :  ||:| || |       ||    |:    :
Db       1171 KEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVAST 1223

Qy        429 WYRHNYLVPTVKQGNYASLLESTSTH---WGFV 458
              ||              |  : : |:::   | |:
Db       1224 WY------------YTHMRDHTNSNGCFWNFI 1243
```

RESULT 8

```
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match             22.5%;  Score 558;  DB 2;  Length 1251;
  Best Local Similarity   31.8%;  Pred. No. 1.6e-24;
  Matches  144;  Conservative   86;  Mismatches  157;  Indels   66;  Gaps   17;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    |::|  |  | :|::: : :  ||||||:: ::  |||   |     |  |
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 892

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
              :  :| ::||: |:|:: |::||||||:||    || |:    ||||: :: ||||| :
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 952

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :  | :::||: |    | : |:|  :|    | ||| |||:||: :|: |:|||| ||
Db        953 SLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLI 1012

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:|      |  |           || | || |||| :|  |:
Db       1013 DKKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1063

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDF 313
              |:    |::||:||| | |:||||::|:    :::||    :|   : | |      |
Db       1064 NNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRL 1121

Qy        314 NEGYKIIIKRI-RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 372
              | |: |:|:   :||| ||  | :| :  : :| : |      :||  :|      |
Db       1122 YSGIKVKIQRVNNSSTNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTNKEKTI 1175

Qy        373 GLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGD 428
              :       |  ::       |:    :  ||:| || |      ||   |:    :
Db       1176 KISSSGNRFNQVVVM------NSVGNCTMNFKNN-NGNN------IGLLGFKADTVVAST 1222

Qy        429 WYRHNYLVPTVKQGNYASLLESTSTH---WGFV 458
              ||            |  : ::|:::   | |:
Db       1223 WY------------YTHMRDNTNSNGFFWNFI 1242


RESULT 9
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxir
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin componente
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
```

C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 22.5%; Score 557.5; DB 1; Length 1296;
Best Local Similarity 31.4%; Pred. No. 1.8e-24;
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

```
Qy          28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGS 84
               |     ||  || :: ||:|:   | |:| | | ::::    |  :||       | | |
Db         862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy          85 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 141
               |     |: |     |||||||:|| |||||| |: :::     ||||: ::||||| :
Db         922 S-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 976

Qy         142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :   :::||:  :: :| : | |     |   | |:| |||:||| : | |||||:||
Db         977 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1036

Qy         201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
               |   :  |  |: |  | |:::   ||        :  :||: | :| |||: |:|  |:
Db        1037 DQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLY 1088

Qy         261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR- 307
               ::   : ::||:||: |:|:| |||:|:   |:|:  |:                |  |
Db        1089 DNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNI 1148

Qy         308 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 367
              |:    | | |||:       |  ||  | :| :: : || | |         | | |
Db        1149 YLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQA 1200

Qy         368 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 427
               :  |    :  |: :     :    |     ::  | |    | : |  || ::|
Db        1201 GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---- 1255

Qy         428 DWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                  | :   |    |   :| :|      | |:|| :
Db        1256 -----NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1289
```

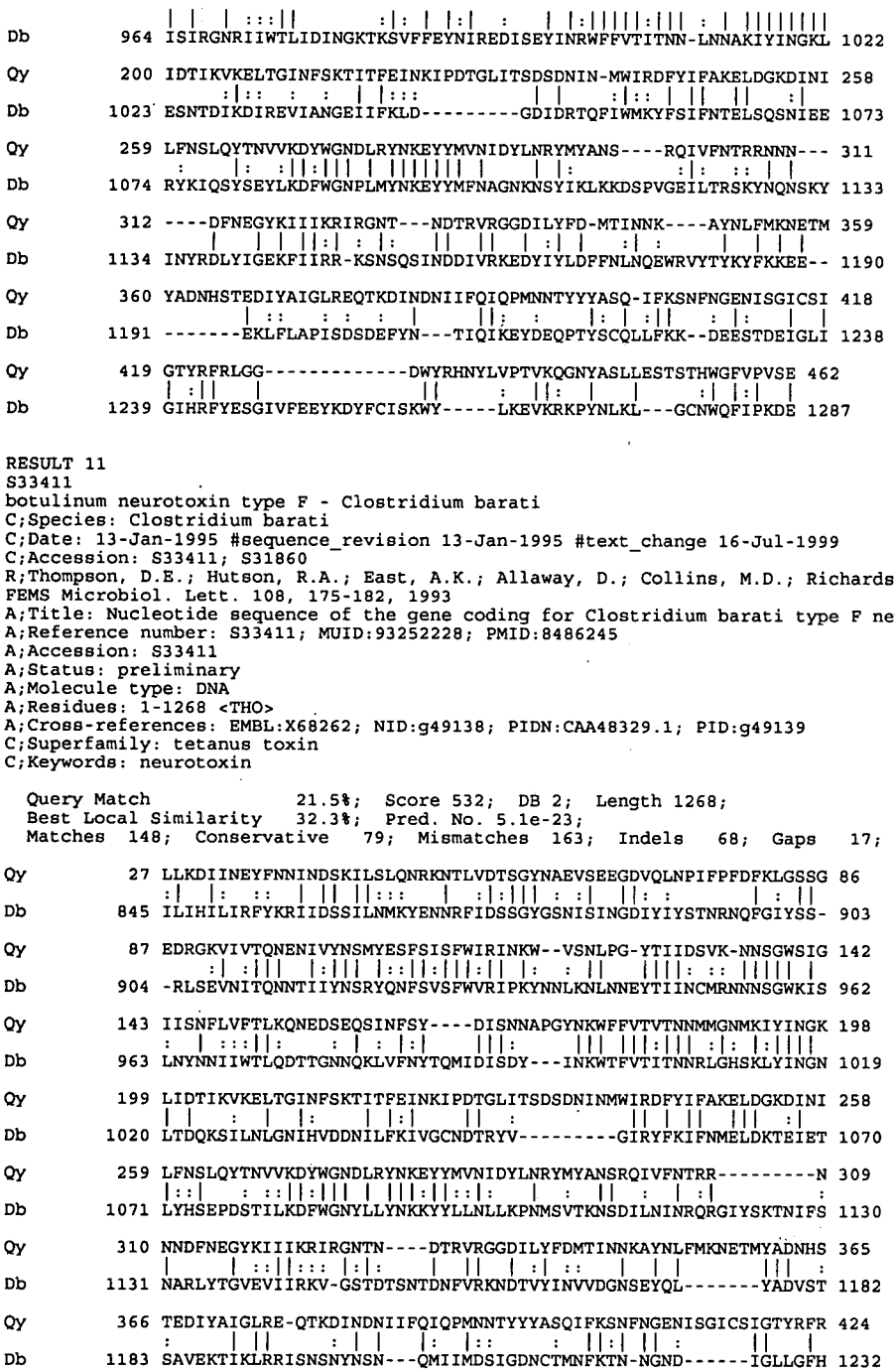
RESULT 10
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisior
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K.
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 22.1%; Score 548; DB 1; Length 1291;
Best Local Similarity 32.7%; Pred. No. 6.3e-24;
Matches 156; Conservative 77; Mismatches 166; Indels 78; Gaps 21;

```
Qy          27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
               :| :: |:| : | :: ||:|: : | |:| ||| |:|     |:||     ||| ||
Db         848 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA 905

Qy          87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140
                 |: ||||:||::||::  ||:|||||| |:    : |     ||||: :||||||
Db         906 --NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy         141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
```



RESULT 11
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 21.5%; Score 532; DB 2; Length 1268;
Best Local Similarity 32.3%; Pred. No. 5.1e-23;
Matches 148; Conservative 79; Mismatches 163; Indels 68; Gaps 17;

RESULT 12
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 21.2%; Score 525.5; DB 2; Length 1296;
Best Local Similarity 30.5%; Pred. No. 1.2e-22;
Matches 142; Conservative 79; Mismatches 177; Indels 67; Gaps 12;

RESULT 13
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 20.6%; Score 509.5; DB 2; Length 1274;
Best Local Similarity 31.2%; Pred. No. 1e-21;
Matches 139; Conservative 79; Mismatches 173; Indels 55; Gaps 15;

```
Qy         33 NEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKV 92
              |  :   | || || ::    |  :| ||| : :|  |:| :         |  |        :|
Db        859 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQF--GIYNSRLSEV 916

Qy         93 IVTQNENIVYNSMYESFSISFWIRINKW---VSNLPGYTIIDSV-KNNSGWSIG---IIS 145
              : || :|:||| |::|||||||:|| |    :::   ||||: :  ||||| |    :
Db        917 NIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRD 976

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIK 204
                :::||:    ::::: | |:  |    | ||| |||:||| :|| :||||| ||
Db        977 CEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKS 1036

Qy        205 VKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQ 264
              :  |  |: |  | |:          |     |   : || | :|  |||  :|  |:::
Db       1037 ISNLGDIHVSDNILFK---------IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEP 1087

Qy        265 YTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN--------NNDFNEG 316
                :::|:|||| | |||:||: |:   ::|:  ||  :  | :|          |    ||
Db       1088 DPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEG 1147

Qy        317 YKIIIKR---IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 373
               ::||::    |  :  |  ||  |: | ::       | |        |||  | ::
Db       1148 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL-------YADTKSEKEKII-- 1198

Qy        374 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDW 429
               :| ::||: : ||  |::     :  |::| || |        ||  |     :   |
Db       1199 ---RTSNLNDS-LGQIIVMDSIGNNCTMNFQNN-NGSN------IGLLGFHSNNLVASSW 1247

Qy        430 YRHNYLVPTVKQGNYASLLESTSTHW 455
              | :|    |   | : | : |    |
Db       1248 YYNNIRRNTSSNGCFWSSI-SKENGW 1272
```

RESULT 14
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglios
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized,
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of syn
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapto
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 16.8%; Score 417; DB 1; Length 1315;
Best Local Similarity 26.5%; Pred. No. 2.3e-16;
Matches 127; Conservative 86; Mismatches 162; Indels 104; Gaps 19;

```
Qy         39 INDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGK 91
              :  | ||:|    : : | ||:|: |    | || |         | :|:       :  :
Db        881 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSE 932

Qy         92 VIVTQNENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWS 140
              ||| :  :| || |: :|::|||:|: |    |:|       |:|| |:| :      ||||
Db        933 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 992

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGK 198
              : :  | |::||| :    : | |  |: :     |  ||| |:|:||: : :  :||||
Db        993 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 1051

Qy        199 LIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINI 258
              |: : ::  |  |     || ::::         :::|  : |  | || | |: |:|
```

```
Db       1052 LMGSAEITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEK 1103

Qy        259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNI-------------DYL---NRYMYANSRQI 302
              |: |      ::|:||| |||: |||:: :             ||:   |   | | :
Db       1104 LYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLN 1163

Qy        303 VFNTRRNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLF 353
              ::  |  |     | | ||||   |   |: |: || :   :: ||  :         | |
Db       1164 IYYRRLYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAF 1218

Qy        354 MKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 413
                : :    ::   |      | |            : :   |: |:    :: :| |
Db       1219 NNLDRILRVGYNAPGIPLYKKMEAVK-----------LRDLKTYSVQL--KLYDDKNAS 1264

Qy        414 GICSIGTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
               :  :||:  ::| |          ||  |:|   :                | ||| |
Db       1265 -LGLVGTHNGQIGNDPNRDILIASNWY-FNHLKDKI----------LGCDWYFVPTDE 1310


RESULT 15
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             9.9%;  Score 245.5;  DB 2;  Length 367;
  Best Local Similarity   34.9%;  Pred. No. 3.4e-07;
  Matches   52;  Conservative   37;  Mismatches   53;  Indels    7;  Gaps    3;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    |::|  |  | :|::: : :  ||||||:: ::  |||   |     |  |
Db        220 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 277

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
              :   :| ::||: |:|:: |::||||||:||    || |:    ||||: :: ||||| :
Db        278 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 337

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISN 170
               :  | :::||: |    | : |:|  :|
Db        338 SLNHNEIIWTLQDNAGINQKLAFNYGNAN 366


Search completed: September 10, 2004, 03:44:14
Job time : 10.1619 secs
```

This Page Blank (usp...

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 5.63701 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 2479
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2317	93.5	1290	1	BXC1_CLOBO	P18640 clostridium
2	801.5	32.3	1276	1	BXD_CLOBO	P19321 clostridium
3	573.5	23.1	1296	1	BXG_CLOBO	Q60393 clostridium
4	558	22.5	1250	1	BXE_CLOBU	P30995 clostridium
5	557.5	22.5	1295	1	BXA1_CLOBO	P10845 clostridium
6	548	22.1	1290	1	BXB_CLOBO	P10844 clostridium
7	547	22.1	1250	1	BXE_CLOBO	Q00496 clostridium
8	525.5	21.2	1295	1	BXA2_CLOBO	Q45894 clostridium
9	509.5	20.6	1274	1	BXF_CLOBO	P30996 clostridium
10	417	16.8	1314	1	TETX_CLOTE	P04958 clostridium
11	222	9.0	1162	1	BXEN_CLOBO	P46082 clostridium
12	222	9.0	1162	1	BXEN_CLOBU	Q06366 clostridium
13	196	7.9	1196	1	BXCN_CLOBO	P46081 clostridium
14	161.5	6.5	396	1	RMAR_YEAST	P02381 saccharomyc
15	142	5.7	1138	1	C7AB_BACUK	Q45708 bacillus th
16	141.5	5.7	476	1	RF3_SACBA	P05512 saccharomyc
17	141.5	5.7	655	1	YKDA_MYCCA	P45615 mycoplasma
18	141	5.7	2136	1	YCF2_MARPO	P09975 marchantia
19	140	5.6	1076	1	RPOB_ASTLO	P27059 astasia lon
20	139.5	5.6	451	1	ARP2_PLAFA	P13824 plasmodium
21	138.5	5.6	1177	1	Y307_MYCGE	P47549 mycoplasma
22	138.5	5.6	2366	1	TOXB_CLODI	P18177 clostridium
23	137.5	5.5	2894	1	YD96_METJA	Q58791 methanococc
24	136	5.5	339	1	RMAR_CANGA	P21358 candida gla
25	135	5.4	537	1	ARP_PLAFA	P04931 plasmodium
26	135	5.4	735	1	CIGB_DICDI	Q94481 dictyosteli
27	134	5.4	559	1	ENS2_YEAST	P12294 saccharomyc
28	134	5.4	760	1	ETF2_MSEPV	Q9yw29 melanoplus
29	133.5	5.4	692	1	Y650_METJA	Q57852 methanococc
30	131.5	5.3	987	1	YD94_METJA	Q58789 methanococc
31	131.5	5.3	1024	1	RPOB_PLAFA	P21421 plasmodium
32	130.5	5.3	643	1	Y903_METJA	Q58313 methanococc
33	130.5	5.3	828	1	PMFC_PROMI	P53514 proteus mir
34	130	5.2	782	1	Y044_UREPA	Q9pra1 ureaplasma
35	130	5.2	1462	1	GTFD_STRMU	P49331 streptococc
36	129	5.2	2710	1	TOXA_CLODI	P16154 clostridium
37	127	5.1	917	1	HXA3_HAEIN	P45355 haemophilus
38	127	5.1	1138	1	C7AA_BACTU	Q03749 bacillus th
39	126.5	5.1	466	1	SYN_BORBU	O51128 borrelia bu
40	126.5	5.1	575	1	RPOC_PLAFA	P21422 plasmodium
41	126.5	5.1	1104	1	COLA_CLOPE	P43153 clostridium
42	126.5	5.1	1694	1	IGA0_HAEIN	P44969 haemophilus
43	126.5	5.1	1702	1	IGA2_HAEIN	P45384 haemophilus
44	126	5.1	968	1	RNE_BUCAP	Q8k9j9 buchnera ap
45	125.5	5.1	3135	1	S230_PLAFO	Q08372 plasmodium

ALIGNMENTS

RESULT 1
BXC1_CLOBO
ID BXC1_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE (Bontoxilysin C1).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90370487; PubMed=2204031;
RA Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA Boquet P., Popoff M.R.;
RT "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL Nucleic Acids Res. 18:4924-4924(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=91024998; PubMed=2222445;
RA Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT "The complete nucleotide sequence of the gene coding for botulinum
RT type C1 toxin in the C-ST phage genome.";
RL Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN [3]
RP SEQUENCE OF 2-25.
RC STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=88153072; PubMed=2450068;
RA Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA Oguma K.;
RT "Establishment of a monoclonal antibody recognizing an antigenic site
RT common to Clostridium botulinum type B, C1, D, and E toxins and
RT tetanus toxin.";
RL Infect. Immun. 56:898-902(1988).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94038966; PubMed=7901002;
RA Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT cleaving HPC-1/syntaxin.";
RL EMBO J. 12:4821-4828(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             93.5%;  Score 2317;  DB 1;  Length 1290;
  Best Local Similarity   99.8%;  Pred. No. 3.5e-130;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        854 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 913

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        914 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 973

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        974 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1033

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1034 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1093

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1094 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1153

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1154 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1213

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1214 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1273

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1274 SLLESTSTHWGFVPVSE 1290


RESULT 2
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
```

```
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match           32.3%;  Score 801.5;  DB 1;  Length 1276;
  Best Local Similarity  41.7%;  Pred. No. 3e-40;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||:||||||||||:|| |||||||||||    :|||| |:  |||| ||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        911 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 967

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
              :  : : |:       :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        968 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 1027

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
              |:::|  :   ||| | |::  |        :|  :||||| ||:|||  :||||::
Db       1028 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 1079

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
               ||:|||||| |::: |||::| :|::||:   |  :|    : :    |  | ||
Db       1080 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKS 1139

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 380
              :      :|:  || :   |  |:: | :    :|:||      |   :||: |:    :

Db       1140 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 1199

Qy        381 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 439
              || |: : :   | |||| |:|  ||   :  |   :||      ::: |    |
Db       1200 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 1251

Qy        440 KQGNYASLLESTSTHWGFV 458
              || : | |||: | |:
Db       1252 T--NYETKLLSTSSFWKFI 1268


RESULT 3
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
```

```
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             23.1%;  Score 573.5;  DB 1;  Length 1296;
  Best Local Similarity   31.8%;  Pred. No. 1e-26;
  Matches  151;  Conservative   82;  Mismatches  169;  Indels   73;  Gaps   16;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|  : | | :||: : |||| |   |:|:||| | ::   ||  | |    ||| :|
Db        852 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNS- 910

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSGWS 140
               :  :   |:: :||:||:::|||:||:|  |: :|          ||||  :||:|||
Db        911 -ENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 969

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
              : |  | :::||        :|| | | | :|   | |||| :|:||: :||  ||||| |
Db        970 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1029

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINIL 259
              : |:  |  || |  | |::     ||           :||:|| || :||:  ::: |
Db       1030 KKSEKILNLDRINSSNDIDFKLINCTDTTKF--------VWIKDFNIFGRELNATEVSSL 1081

Qy        260 FNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE---- 315
              :     || :||:||| |||: :||: |    | |:   |:  :  |    | ||
Db       1082 YWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAAIN 1140

Qy        316 ------GYKIIIKRI---RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST 366
                   | : |||:    |   ||  || || :| ::                  || |
Db       1141 YQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI-----------------DNISD 1182

Qy        367 EDIYAIGLREQTKDINDNIIFQIQPMNN--TYYYASQI----FKSNFNGENIS------- 413
              |  | : :   :|:|   :   : |:|:  |:|   ||      |: :| : :
Db       1183 ES-YRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTF 1239

Qy        414 GICSIGTYRFRLGGDWYRH-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 462
              |:  || :    |  |  : ||    : |     : |: :      :| |:|| |
Db       1240 GLFGIGKFVKDYGYVWDTYDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 1292


RESULT 4
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             22.5%;  Score 558;  DB 1;  Length 1250;
  Best Local Similarity   31.8%;  Pred. No. 8.1e-26;
  Matches  144;  Conservative   86;  Mismatches  157;  Indels   66;  Gaps   17;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    :|::|  |  | :|::: : :  ||||||:: ::  |||   |     |  |
Db        834 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 891

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
              :   :| ::||: |:|:: |::||||||:||    || |:    ||||: :: ||||| :
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 951
```

```
Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :  | :::||: |    | : |:|  :|    | ||| |||:||: :|: |:|||| ||
Db        952 SLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLI 1011

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:|      |  |           || | || ||||  :|  |:
Db       1012 DKKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1062

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDF 313
              |:    |::||:||| | |:|||::|:    :::||    :|   : | |      |
Db       1063 NNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRL 1120

Qy        314 NEGYKIIIKRI-RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 372
              | |: |:|:   :|||  ||  | :| :  : :| : |      :|||  :|     |
Db       1121 YSGIKVKIQRVNNSSTNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTNKEKTI 1174

Qy        373 GLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGD 428
              :       |  ::       |:    :  ||:| || |      ||   |:    :
Db       1175 KISSSGNRFNQVVVM------NSVGNCTMNFKNN-NGNN------IGLLGFKADTVVAST 1221

Qy        429 WYRHNYLVPTVKQGNYASLLESTSTH---WGFV 458
              ||              |  : ::|:::   | |:
Db       1222 WY------------YTHMRDNTNSNGFFWNFI 1241
```

RESULT 5
BXA1_CLOBO
ID BXA1_CLOBO STANDARD; PRT; 1295 AA.
AC P10845; P01561; P18639;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 2916;
RX MEDLINE=90235864; PubMed=2185020;
RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type A
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 189:73-81(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=62A;
RX MEDLINE=90264400; PubMed=2160960;
RA Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [3]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=62A;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN=Hall;
RX MEDLINE=89350959; PubMed=2669749;
RA Betley M.J., Somers E., Dasgupta B.R.;
RT "Characterization of botulinum type A neurotoxin gene: delineation of
RT the N-terminal encoding region.";
RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN [5]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=Type A NIH;
RX MEDLINE=96096783; PubMed=8521962;
RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT components of Clostridium botulinum type A progenitor toxins.";
RL FEBS Lett. 376:41-44(1995).
RN [6]
RP SEQUENCE OF 1-16.
RX MEDLINE=84178501; PubMed=6370252;
RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequence of the heavy and light chains of
RT botulinum neurotoxin type A.";
RL Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN [7]
RP SEQUENCE OF 1-46.
RA Dasgupta B.R., Foley J., Niece R.;
RT "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL Biochemistry 26:4162-4162(1987).
RN [8]
RP SEQUENCE OF 1-5 AND 444-456.
RX MEDLINE=91120847; PubMed=2126206;
RA Dasgupta B.R., Dekleva M.L.;
RT "Botulinum neurotoxin type A: sequence of amino acids at the
RT N-terminus and around the nicking site.";
RL Biochimie 72:661-664(1990).
RN [9]
RP SEQUENCE OF 448-464 AND 872-895.
RX MEDLINE=89024662; PubMed=3178218;
RA Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT halves and their partial sequences.";
RL Arch. Biochem. Biophys. 266:142-151(1988).
RN [10]
RP SEQUENCE OF 448-482.
RX MEDLINE=85285016; PubMed=3896784;
RA Shone C.C., Hambleton P., Melling J.;
RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT and purification of two tryptic fragments. Proteolytic action near
RT the COOH-terminus of the heavy subunit destroys toxin-binding
RT activity.";
RL Eur. J. Biochem. 151:75-82(1985).
RN [11]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA Benfenati F., Wilson M.C., Montecucco C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
RN [12]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;
RT "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
RN [13]
RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX MEDLINE=21556941; PubMed=11700044;
RA Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT "Site-directed mutagenesis identifies active-site residues of the
RT light chain of botulinum neurotoxin type a.";
RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX MEDLINE=98455071; PubMed=9783750;
RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;

```
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             22.5%;  Score 557.5;  DB 1;  Length 1295;
  Best Local Similarity   31.4%;  Pred. No. 9.1e-26;
  Matches  144;  Conservative   80;  Mismatches  180;  Indels   55;  Gaps   11;

Qy         28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFD---FKLGS 84
              |     ||  || :: ||:|:   | |:| | | ::::    |  :||       | | |
Db        861 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 920

Qy         85 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 141
              |     |: |    ||||||||:|| |||||| |: :::     ||||: ::||||| :
Db        921 S-----KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 975

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :    :::||:  :: :| : | |     |   | |:| |||:||| : | |||||:||
Db        976 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 1035

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:::   ||          :  :||: | :| |||: |:|  |:
Db       1036 DQKPISNLGNIHASNNIMFKLDGCRDT--------HRYIWIKYFNLFDKELNEKEIKDLY 1087

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS------------RQIVFNTR- 307
              ::   : ::||:||: |:|:| |||:|:   |:|:  |:                |  |
Db       1088 DNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNI 1147

Qy        308 RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTE 367
                |:    | | |||:      |  ||  | :| :: : || | |         | | |
Db       1148 YLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRL--------ATNASQA 1199

Qy        368 DIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGG 427
              : |    :  |: :    :   |    ::  | |    | : |  || ::|
Db       1200 GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGFHQF---- 1254

Qy        428 DWYRHNYLVPTVKQGNYASLLESTS----THWGFVPVSE 462
                | :   |    |   :| :|      | |:|| :
Db       1255 -----NNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDD 1288


RESULT 6
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
```

```
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             22.1%;  Score 548;  DB 1;  Length 1290;
  Best Local Similarity   32.7%;  Pred. No. 3.3e-25;
  Matches  156;  Conservative   77;  Mismatches  166;  Indels   78;  Gaps   21;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :| :: |:| : | :: ||:|: : | |:| ||| |:|      |:||      ||| ||
Db        847 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA 904

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140
                 |: ||||:||::||::  ||:|||||| |:    : |     ||||: :||||||
Db        905 --NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 962

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
              | |  | :::||        :|: | |:|  :   | |:|||||:||| : | ||||||||
Db        963 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 1021

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINI 258
                :|::  :  :  | |:::          | |    :|:: | ||  ||   :|
Db       1022 ESNTDIKDIREVIANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIEE 1072

Qy        259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN--- 311
               :   |:  :||:||| | ||||||| |    | |:          :|:  :: | |
Db       1073 RYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKY 1132

Qy        312 ----DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFD-MTINNK----AYNLFMKNETM 359
                   | | ||:| : |:   ||  ||  | :| |   :| :     | | | |
Db       1133 INYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE-- 1189

Qy        360 YADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQ-IFKSNFNGENISGICSI 418
                 | ::   : :  :   |     ||:  :    |: | :||   : |:   |  |
Db       1190 -------EKLFLAPISDSDEFYN---TIQIKEYDEQPTYSCQLLFKK--DEESTDEIGLI 1237

Qy        419 GTYRFRLGG-------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
               | :||  |              ||     :  ||:  |   |     :| |:|  |
Db       1238 GIHRFYESGIVFEEYKDYFCISKWY-----LKEVKRKPYNLKL---GCNWQFIPKDE 1286


RESULT 7
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
```

```
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             22.1%;  Score 547;  DB 1;  Length 1250;
  Best Local Similarity   31.5%;  Pred. No. 3.7e-25;
  Matches  142;  Conservative   84;  Mismatches  163;  Indels   62;  Gaps   16;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    |::|  |  | :|::: : :  ||||||:: ::  |||   |     |  |
Db        834 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 891

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
              :  :| ::||: |:|:: |::||||||:||    || |:    ||||: :: ||||| :
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 951

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :  | :::| : |    | : |:|  :|    | ||| |||:||: :|: |:|||| ||
Db        952 SLNHNEIIWTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLI 1011

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:|     |  |          || | || |||| :|  |:
Db       1012 DQKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1062

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYM-YANSRQIVFNTRRN---NNDFNE 315
              ::  ||::||:||| | |:||||::|:   | ::    :|   : | |      |
Db       1063 SNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYS 1122
```

```
Qy        316 GYKIIIKRI-RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGL 374
              | |: |:|:   :|||  ||  | :| :   :    :||     :|||  :|      | :
Db       1123 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKT--HLF----PLYADTATTNKEKTIKI 1176

Qy        375 REQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDWY 430
                  |  ::       |:    :  ||:| || |       ||   |:    :   ||
Db       1177 SSSGNRFNQVVVM------NSVGNCTMNFKNN-NGNN------IGLLGFKADTVVASTWY 1223

Qy        431 RHNYLVPTVKQGNYASLLESTSTH---WGFV 458
                      |  : : |:::   | |:
Db       1224 ------------YTHMRDHTNSNGCFWNFI 1241


RESULT 8
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             21.2%;  Score 525.5;  DB 1;  Length 1295;
  Best Local Similarity   30.5%;  Pred. No. 7.2e-24;
  Matches  142;  Conservative   79;  Mismatches  177;  Indels   67;  Gaps   12;

Qy         28 LKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE 87
              |     ||  || :: |||:  :|: |:| | | |:::    |  : |      || :
Db        861 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 920

Qy         88 DRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNL---PGYTIIDSVKNNSGWSIGII 144
                :||:     |||||||:|| ||||:| |: | :     ||||: ::||||| : :
Db        921 STIEVIL--KNAIVYNSMYENFSTSFWIKIPKYFSKINLNNEYTIINCIENNSGWKVSLN 978

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
                :::||: |: : | : | |     |   | |:| |||:||| :   ||||||:|||
Db        979 YGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQK 1038

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
               :  |  |: |  | |:::   |              : |: | :| |||: |:|  |::|
Db       1039 PISNLGNIHASNKIMFKLDGCRD--------PRRYIMIKYFNLFDKELNEKEIKDLYDSQ 1090

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS-------------RQIVFNTRRNN 310
               : ::||:||| |:|:| |||:|:   |:|:  |:               :|      |
Db       1091 SNSGILKDFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLN 1150

Qy        311 NDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIY 370
               :  || | |||:       |  ||  | :| :: : || | |          | | |   :
Db       1151 STLYEGTKFIIKKYASGNEDNIVRNNDRVYINVVVKNKEYRL--------ATNASQAGVE 1202

Qy        371 AIGLREQTKDINDNIIFQI--------QPMNNTYYYASQIFKSNFNGENISGICSIGTYR 422
               |    :  |: :  : |:         | : |          | |    | : |  || :
Db     -  1203 KILSALEIPDVGN--LSQVVVMKSKDDQGIRNK-------CKMNLQDNNGNDIGFIGFHL 1253

Qy        423 F-----RLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
               :      :  :||         :|   ||   :    | |:|| :
Db       1254 YDNIAKLVASNWYN--------RQVGKAS--RTFGCSWEFIPVDD 1288


RESULT 9
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
```

```
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match           20.6%;  Score 509.5;  DB 1;  Length 1274;
  Best Local Similarity  31.2%;  Pred. No. 6.3e-23;
  Matches  139;  Conservative   79;  Mismatches  173;  Indels   55;  Gaps   15;

Qy         33 NEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKV 92
              |  :  | || || ::    |  :| ||| : :|  |:| :          |  |     :|
Db        859 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQF--GIYNSRLSEV 916

Qy         93 IVTQNENIVYNSMYESFSISFWIRINKW---VSNLPGYTIIDSV-KNNSGWSIG---IIS 145
                : || :|:||| |::|||||:|| |    :::   ||||: :  |||| |    :
Db        917 NIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRD 976

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIK 204
              :::||:    ::::: | |:  |     | ||| |||:||| :|| :|||| ||
Db        977 CEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKS 1036

Qy        205 VKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQ 264
              :  |  |: |  | |:          |    |  : || | :|  ||| :|  |:::
Db       1037 ISNLGDIHVSDNILFK---------IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEP 1087

Qy        265 YTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN--------NNDFNEG 316
                :::|:|||| | |||:||: |:   ::|:  ||  :  | :|          |    ||
Db       1088 DPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEG 1147

Qy        317 YKIIIKR---IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 373
                ::||::   |  :  |  ||  |: | ::       | |        |||  | ::
Db       1148 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL-------YADTKSEKEKII-- 1198

Qy        374 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGDW 429
                :| ::||: : ||  |::     :  |::| || |      ||   |    :   |
Db       1199 ---RTSNLNDS-LGQIIVMDSIGNNCTMNFQNN-NGSN------IGLLGFHSNNLVASSW 1247

Qy        430 YRHNYLVPTVKQGNYASLLESTSTHW 455
               | :|    |   | : | : |    |
Db       1248 YYNNIRRNTSSNGCFWSSI-SKENGW 1272


RESULT 10
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
```

```
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).

CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
```

```
FT   TURN         969    970
FT   STRAND       972    977
FT   STRAND       980    981
FT   HELIX        983    985
FT   STRAND       987    995
FT   TURN         996    997
FT   STRAND       998   1004
FT   TURN        1006   1007
FT   STRAND      1010   1016
FT   STRAND      1020   1020
FT   TURN        1021   1022
FT   STRAND      1031   1037
FT   TURN        1039   1040
FT   STRAND      1042   1047
FT   TURN        1048   1049
FT   STRAND      1050   1056
FT   TURN        1058   1059
FT   STRAND      1068   1074
FT   TURN        1079   1080
FT   STRAND      1082   1091
FT   HELIX       1097   1105
FT   TURN        1106   1107
FT   STRAND      1112   1112
FT   STRAND      1114   1114
FT   TURN        1116   1117
FT   STRAND      1120   1120
FT   STRAND      1122   1122
FT   TURN        1123   1124
FT   STRAND      1127   1131
FT   HELIX       1132   1134
FT   TURN        1135   1136
FT   STRAND      1137   1141
FT   TURN        1144   1145
FT   STRAND      1148   1152
FT   STRAND      1155   1158
FT   TURN        1159   1162
FT   STRAND      1163   1166
FT   STRAND      1173   1178
FT   TURN        1184   1185
FT   STRAND      1188   1188
FT   STRAND      1190   1190
```

Query Match 16.8%; Score 417; DB 1; Length 1314;
Best Local Similarity 26.5%; Pred. No. 2e-17;
Matches 127; Conservative 86; Mismatches 162; Indels 104; Gaps 19;

```
Qy        39 INDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGE-------DRGK 91
Db       880 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVP--------GINGKAIHLVNNESSE 931

Qy        92 VIVTQNENIVYNSMYESFSISFWIRINK-WVSNLP-----GYTIIDSVKNN-----SGWS 140
Db       932 VIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWS 991

Qy       141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY--NKWFFVTVTNNMMGNMKIYINGK 198
Db       992 VSLKGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGV 1050

Qy       199 LIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINI 258
Db      1051 LMGSAEITGLGAIREDNNITLKLDR--------CNNNNQYVSIDKFRIFCKALNPKEIEK 1102

Qy       259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNI-------------DYL---NRYMYANSRQI 302
Db      1103 LYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLN 1162

Qy       303 VFNTRRNNNDFNEGYKIIIKRIRGNTN-DTRVRGGDILYFDMTINNKAY--------NLF 353
Db      1163 IYYRRLYN-----GLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAF 1217

Qy       354 MKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENIS 413
Db      1218 NNLDRILRVGYNAPGIPLYKKMEAVK------------LRDLKTYSVQL--KLYDDKNAS 1263
```

```
Qy       414 GICSIGTYRFRLGGD----------WYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
Db      1264 -LGLVGTHNGQIGNDPNRDILIASNWY-FNHLKDKI-----------LGCDWYFVPTDE 1309
```

RESULT 11
BXEN_CLOBO
ID BXEN_CLOBO STANDARD; PRT; 1162 AA.
AC P46082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
GN ENT-120.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mashike;
RX MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
RT component of Clostridium botulinum type E progenitor toxin.";
RL J. Gen. Microbiol. 139:79-86(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D12697; BAA02194.1; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDDAE0F39D CRC64;

Query Match 9.0%; Score 222; DB 1; Length 1162;
Best Local Similarity 22.4%; Pred. No. 6.5e-06;
Matches 98; Conservative 68; Mismatches 153; Indels 118; Gaps 17;

```
Qy        31 IINEYFNNINDSKILSLQNRK----------------------------NTLVDTSGY 60
Db       763 IMQNSFSNL-DFDFLDIQNMKNLFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGK 821

Qy        61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVY------NSMYESFSISFW 114
Db       822 NTLVE----------YPKDIGLVYGINNNAIHLTGANQNIKFTNDYFENGLTNNFSIYFW 871

Qy       115 IRINKWVSNLPGYTI----IDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISN 170
Db       872 LR------NLKQNTIKSKLIGSKEDNCGWEIYFENDGLVFNIIDSNGNEKNIYLS-NISN 924

Qy       171 NAPGYNKWFFVTVT-NNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGL 229
Db       925 KS-----WHYIVISINRLKDQLLIFIDNILVANEDIKEILNIYSS-------------DI 966

Qy       230 ITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNID 289
Db       967 ISLLSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNRTYELFNYV 1026

Qy       290 Y----LNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTR------------V 333
```

```
Db       1027 FPEIAINKIEQNNNIYLSINNENNLNFKPLKFKLL------NTNPNKQYVQKWDEVIFSV 1080

Qy        334 RGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPM- 392

Db       1081 LDGTEKYLDISTTN--------NRIQLVDNKNNAQIFIIN--------NDIFISNCLTLT 1124

Qy        393 -NNTYYYASQIFKSNFN 408

Db       1125 YNNVNVYLS-IKNQDYN 1140


RESULT 12
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match             9.0%;  Score 222;  DB 1;  Length 1162;
  Best Local Similarity   22.5%;  Pred. No. 6.5e-06;
  Matches   99;  Conservative   70;  Mismatches  147;  Indels  124;  Gaps   19;

Qy         31 IINEYFNNINDSKILSLQNRK----------------------------NTLVDTSGY 60

Db        763 IMQNSFSNL-DFDFLDIQNMKNLFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGK 821

Qy         61 NAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVY------NSMYESFSISFW 114

Db        822 NTLVE----------YPKDIGLVYGINNNAIHLTGANQNIKFTNDYFENGLTNNFSIYFW 871

Qy        115 IRINKWVSNLPGYTI----IDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISN 170

Db        872 LR------NLNQNTIKSKLIGSKEDNCGWEIYFENNGLVFNIIDSNGNEKNIYLS-NISN 924

Qy        171 NAPGYNKWFFVTVT-NNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGL 229

Db        925 KS-----WHYIVISINRLKDQLLIFIDNILVANEDIKEILNIYSS------------DI 966

Qy        230 ITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNID 289

Db        967 ISLLSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNRTYELFNYV 1026

Qy        290 Y----LNRYMYANSRQIVFNTRRNNNDFN---EGYKIIIKRIRGNTNDTR---------- 332

Db       1027 FPEIAINKIEQNNN---IYLSNNNENSLNFKPLKFKLL------NTNPNKQYVQKWDEVI 1077

Qy        333 --VRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQ 390

Db       1078 FSVLDGTEKYLDISIDN--------NRIQLVDNKNNAKTFIIN--------NDIFISNCL 1121

Qy        391 PM--NNTYYYASQIFKSNFN 408

Db       1122 TLTYNNVNVYLS-IKNQDYN 1140


RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             7.9%;  Score 196;  DB 1;  Length 1196;
  Best Local Similarity   18.5%;  Pred. No. 0.00023;
  Matches   87;  Conservative   85;  Mismatches  149;  Indels  150;  Gaps   17;

Qy         31 IINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRG 90

Db        797 LINSYTFKTIDFKFLDIQSIKNF------FNSQVEQ---VMKEILSPYQLLLFASKGPNS 847

Qy         91 KVI-----------VTQNENIVY----------------------NSMYESFSISFWI 115

Db        848 NIIEDISGKNTLIQYTESIELVYGVNGESLYLKSPNETIKFSNKFFTNGLTNNFTICFWL 907

Qy        116 RINKWVSNLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY 175
```

```
Db        908 RFTG--KNDDKTRLIGNKVNNCGWEIYFEDNGLVFEIIDSNGNQESVYLSNIINDN---- 961

Qy        176 NKWFFVTVT-NNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDS 234

Db        962 --WYYISISVDRLKDQLLIFINDKNVANVSIDQILSI-YSTNIISLVNK---------- 1007

Qy        235 DNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRY 294

Db       1008 -NNSIYVEELSVLDNPITSEEVIRNYFSYLDNSYIRDSSKSLLEYNKNYQLYNYVFPETS 1066

Qy        295 MYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFM 354

Db       1067 LY----------------------------EVND--------------NNKSY-LSL 1080

Qy        355 KNETMYADNHSTEDIYAIGLREQTKDINDNIIF------------------QIQPMNNT 395

Db       1081 KN---------TDGINISSVKFKLINIDESKVYVQKWDECIICVLDGTEKYLDISPENNR 1131

Qy        396 YYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN--YLVPTVKQGNY 444

Db       1132 I----QLVSSKDNAKKI----TVNTDLFRPDCITFSYNDKYFSLSLRDGDY 1174
```

RESULT 14
RMAR_YEAST
ID RMAR_YEAST STANDARD; PRT; 396 AA.
AC P02381;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Mitochondrial ribosomal protein VAR1.
GN VAR1.
OS Saccharomyces cerevisiae (Baker's yeast).
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83050946; PubMed=6754087;
RA Hudspeth M.E.S., Ainley W.M., Shumard D.S., Butow R.A., Grossman L.I.;
RT "Location and structure of the var1 gene on yeast mitochondrial DNA:
RT nucleotide sequence of the 40.0 allele.";
RL Cell 30:617-626(1982).
CC -!- FUNCTION: Essential for mitochondrial protein synthesis and
CC required for the maturation of small ribosomal subunits.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: BELONGS TO THE VAR1 FAMILY OF RIBOSOMAL PROTEINS.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; V00705; CAA24077.1; -.
DR GermOnline; 144534; -.
DR SGD; S0007275; VAR1.
DR InterPro; IPR007980; Yeast_VAR1.
DR Pfam; PF05316; Yeast_VAR1; 1.
KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 396 AA; 46881 MW; 62A68A24740BCBE0 CRC64;

Query Match 6.5%; Score 161.5; DB 1; Length 396;
Best Local Similarity 22.6%; Pred. No. 0.0072;
Matches 97; Conservative 80; Mismatches 156; Indels 97; Gaps 23;

```
Qy         27 LLKDIINEYFNNIN---DSKILSLQNR----KNTLVDTSGYNAEVSEE--GDVQLNP--I 75

Db          8 MILSMMNKTNNNNNIIINNTLDSLMNKKLLLKNMLLDMNNKKMNNMKRMLNNNNMNPAGA 67
```

```
Qy         76 FPFDFKLGSSGEDRGKV-----------IVTQNENI-VYNSMYESFSISFWIRINKWVS 122

Db         68 NPVVHRIGPAGNINNKLQHLNNMNNWNTQIYNYNKNMEIMNTMNDKL-------INKLL- 119

Qy        123 NLPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSY---DISNNAPGYNKWF 179

Db        120 ----YKMMTLKLNNMNINKIIMSK----TINQHSLNKLNIKFYYYNNDINNN--NNNNNY 169

Qy        180 FVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINM 239

Db        170 YMNMMNKLMNIMNNNNMNNNLCN------ILSYYYNKKVTIEPIKL---SYIYLNSD---- 216

Qy        240 WIRDFYIFAKELDGKDINILFN-SLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYAN 298

Db        217 -IFSKYISLNDMDKYNNGILTNYQRMLNNIMPKLNDHNISMN---YINNINNINNNKYNN 272

Qy        299 SRQIVFNTR--RNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFM-- 354

Db        273 MINLLNNNNNINNNNNYN-----------NNNNNYIGNINNIYNNMTIDNIPMDILMYK 320

Qy        355 -------KNETMYADNH---STEDIYAIGLREQTK----DINDNIIFQIQPMNNTYYYAS 400

Db        321 YLVGWSIKFKGRLSNNNGRTSTTNLLN-GTFNNKKYLWSNINNNYKLNYIPSNHNLYNNS 379

Qy        401 QIFKSNFNGE 410

Db        380 NI---NKNGK 386
```

RESULT 15
C7AB_BACUK
ID C7AB_BACUK STANDARD; PRT; 1138 AA.
AC Q45708;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidial crystal protein cry7Ab (Insecticidal delta-endotoxin
DE CryVIIA(b)) (Crystaline entomocidal protoxin) (130 kDa crystal
DE protein).
GN CRY7AB OR CRYVIIA(B).
OS Bacillus thuringiensis (subsp. kumamotoensis).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=132267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD867;
RA Payne J.M., Fu J.M.;
RT "Coleopteran-active Bacillus thuringiensis isolates and genes encoding
RT coleopteran-active toxins.";
RL Patent number US5286486, 15-FEB-1994.
CC -!- FUNCTION: PROMOTES COLLOIDOSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF COLEOPTERA.
CC -!- DEVELOPMENTAL STAGE: The crystal protein is produced during
CC sporulation and is accumulated both as an inclusion and as part of
CC the spore coat.
CC -!- MISCELLANEOUS: Toxic segment of the protein is located in the N-
CC terminus.
CC -!- SIMILARITY: Belongs to the delta endotoxin family.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; U04368; AAA21121.1; -.
DR HSSP; P07130; 1DLC.
DR InterPro; IPR001178; Endotoxin.
DR InterPro; IPR005638; endotoxin_C.
DR InterPro; IPR005639; endotoxin_N.

```
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF00555; endotoxin; 1.
DR   Pfam; PF03944; endotoxin_C; 1.
DR   Pfam; PF03945; endotoxin_N; 1.
KW   Toxin; Sporulation.
SQ   SEQUENCE   1138 AA;  129658 MW;  E12DC80C0A56D1DA CRC64;

  Query Match             5.7%;  Score 142;  DB 1;  Length 1138;
  Best Local Similarity   25.0%;  Pred. No. 0.35;
  Matches   90;  Conservative   49;  Mismatches  117;  Indels  104;  Gaps   21;

Qy        125 PGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVT 184
              ||:|  | ||  |   || |   :   | ||          | :              |
Db        518 PGFTGGDLVKRGSNGYIGDIKATVNSPLSQN----------YRVR-----------VRYA 556

Qy        185 NNMMGNMKIYINGKL---------IDTI-KVKELT-----GINFSKTITFEINKIPDTGL 229
               |: |   :||| |:         ::|| : |:||       | :| || |  :| |   |
Db        557 TNVSGQFNVYINDKITLQRKFQNTVETIGEGKDLTYGSFGYIEYSTTIQFP-DKHPKITL 615

Qy        230 ITSDSDNINMWIRDFYIFAKEL-------DGKD--------INILF----NSLQ------ 264
              ||  | :     ||: : |         | |:        :| ||    |:||
Db        616 HLSDLSNNS----SFYVDSIEFIPVDVNYDEKEKLEKAQKAVNTLFTEGRNALQKDVTDY 671

Qy        265 ---YTNVVKDYWGNDLRYNKEYYMVN-IDYLNRYMYANSRQIVFN---TRRNNNDFNEGY 317
                 ::: |    ||  |::  : | : |  |  |  || :: :       |::: |  |
Db        672 KVDQVSILVDCISGDLYPNEKRELQNLVKYAKRLSY--SRNLLLDPTFDSINSSEENGWY 729

Qy        318 ---KIIIKR----IRGN------TNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADN- 363
               |:|        :||       ||||:        :   :  |     :|  | :|
Db        730 GSNGIVIGNGDFVFKGNYLIFSGTNDTQ--------YPTYLYQKIDESKLKEYTRYKLKG 781

Qy        364 --HSTEDIYAIGLR----EQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICS 417
                |::|: |  :|     :| |::||::  | |  ||    ::     :  || |  ||
Db        782 FIESSQDLEAYVIRYDAKHRTLDVSDNLLPDILP-ENTCGEPNRCAAQQYLDENPSSECS 840

Search completed: September 10, 2004, 03:35:47
Job time : 7.63701 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 21.1388 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-62
Perfect score: 2479
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NYASLLESTSTHWGFVPVSE 462

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2317	93.5	1291	2	Q93HT3	Q93ht3 clostridium
2	1755	70.8	1285	2	Q9LBR1	Q9lbr1 clostridium
3	1754	70.8	1285	2	Q45967	Q45967 clostridium
4	814.5	32.9	1280	2	Q841S3	Q841s3 clostridium
5	813.5	32.8	1280	2	Q45849	Q45849 clostridium
6	813.5	32.8	1280	2	Q9LBS7	Q9lbs7 clostridium
7	801.5	32.3	1275	12	Q9QTG7	Q9qtg7 clostridium
8	566.5	22.9	1252	2	Q8KZM3	Q8kzm3 clostridium
9	566.5	22.9	1255	2	Q9FAR6	Q9far6 clostridium
10	565	22.8	1291	2	Q933K0	Q933k0 clostridium
11	565	22.8	1291	2	Q08077	Q08077 clostridium
12	559	22.5	1291	2	Q9ZAJ8	Q9zaj8 clostridium
13	559	22.5	1291	2	Q8GR96	Q8gr96 clostridium
14	558	22.5	1251	2	Q9K395	Q9k395 clostridium
15	557	22.5	1291	2	Q93G71	Q93g71 clostridium
16	550	22.2	441	2	Q9X708	Q9x708 clostridium
17	532	21.5	1268	2	Q45851	Q45851 clostridium
18	525	21.2	1278	2	Q57236	Q57236 clostridium
19	487.5	19.7	1280	2	Q9ZAJ5	Q9zaj5 clostridium
20	425	17.1	451	2	Q9LA13	Q9la13 clostridium
21	420	16.9	1310	2	Q93N27	Q93n27 clostridium
22	271	10.9	361	2	Q45848	Q45848 clostridium
23	271	10.9	361	2	Q45846	Q45846 clostridium
24	245.5	9.9	367	2	Q45861	Q45861 clostridium
25	245.5	9.9	367	2	Q45862	Q45862 clostridium
26	245.5	9.9	1197	2	Q45888	Q45888 clostridium
27	238	9.6	1198	2	O06018	O06018 clostridium
28	229	9.2	1197	2	Q9ZAJ9	Q9zaj9 clostridium
29	229	9.2	1197	2	O33871	O33871 clostridium
30	227	9.2	1198	2	O69277	O69277 clostridium
31	226.5	9.1	1197	2	P71117	P71117 clostridium
32	226	9.1	1198	2	Q45893	Q45893 clostridium
33	213	8.6	1193	2	P71107	P71107 clostridium
34	213	8.6	1193	2	Q45914	Q45914 clostridium
35	212	8.6	1193	2	Q45880	Q45880 clostridium
36	211	8.5	1161	2	O69276	O69276 clostridium
37	211	8.5	1161	2	Q45891	Q45891 clostridium
38	204.5	8.2	1162	2	Q9ZAJ6	Q9zaj6 clostridium
39	197	7.9	1196	2	Q9LBS8	Q9lbs8 clostridium
40	196	7.9	1196	2	Q45916	Q45916 clostridium
41	196	7.9	1196	2	Q93HT4	Q93ht4 clostridium
42	196	7.9	1196	9	Q38197	Q38197 clostridium
43	196	7.9	3063	5	Q8I284	Q8i284 plasmodium
44	195.5	7.9	1165	2	Q45844	Q45844 clostridium
45	195.5	7.9	1165	2	Q45887	Q45887 clostridium

ALIGNMENTS

RESULT 1
Q93HT3
ID Q93HT3 PRELIMINARY; PRT; 1291 AA.
AC Q93HT3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-Yoichi;
RX MEDLINE=21534265; PubMed=11676492;
RA Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,
RA Fujinaga Y., Oguma K., Ohyama T.;
RT "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in
RT Hemagglutination.";
RL Biochem. Biophys. Res. Commun. 288:650-657(2001).
DR EMBL; AB061780; BAB71749.1; -.
DR PIR; A43503; A43503.
DR PIR; S11291; A49777.
DR PIR; S46431; S46431.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  148869 MW;  4A21DB35B8743CF8 CRC64;

  Query Match           93.5%;  Score 2317;  DB 2;  Length 1291;
  Best Local Similarity   99.8%;  Pred. No. 5.6e-116;
  Matches  436;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 974

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 1034

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 1094

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 1154

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1214

Qy        386 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 IFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYA 1274

Qy        446 SLLESTSTHWGFVPVSE 462
              |||||||||||||||||
Db       1275 SLLESTSTHWGFVPVSE 1291


RESULT 2
Q9LBR1
ID   Q9LBR1      PRELIMINARY;      PRT;  1285 AA.
AC   Q9LBR1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-4947;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA   Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT   "Characterization of the Progenitor Toxin Components Produced by
RT   Clostridium botulinum Type D Strain 4947.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037920; BAA90661.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1285 AA;  147352 MW;  B63AFA487D570680 CRC64;

  Query Match           70.8%;  Score 1755;  DB 2;  Length 1285;
  Best Local Similarity   77.0%;  Pred. No. 6.5e-86;
  Matches  339;  Conservative   37;  Mismatches   56;  Indels    8;  Gaps    4;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :||||:||||||:||||||||||:||||:|||||||  ||:||||||||||||||||
Db        851 SLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSS 910

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
              |:||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDDRGKVIVTQNENIVYNAMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 970

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
              ||||||||||:||| |||||||| || |||||||||:| |||||| ||||||||||||
Db        971 NFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFVTITTNMMGNMMIYINGKLIDTIKV 1030

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
              ||||||||||||::||||:|||||||||||||||||||||||| ||||||||||||
Db       1031 KELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSLQY 1090

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
              ||||||||||||||||||:|||||:|::|:||||      ||||||:||||||||||||
Db       1091 TNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFNEGYKIIIKRIR 1150

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAIGLREQTKD-IND 383
              |||||||||| ::|||: ||:|| |:|      ||  : :   |:  :|  :|  | |
Db       1151 GNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYKPSRNLGTDLVPLGALDQPMDEIRK 1205

Qy        384 NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGD-WYRHNYLVPTVKQG 442
                | ||| |   |||||:| |:    |  || |||:| |:|| | |: | ||:|  :|
Db       1206 YGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLGDDYWFNHEYLIPVIKIE 1265

Qy        443 NYASLLESTSTHWGFVPVSE 462
              :||||||||||| ||| ||
Db       1266 HYASLLESTSTHWVFVPASE 1285


RESULT 3
Q45967
ID   Q45967      PRELIMINARY;      PRT;  1285 AA.
AC   Q45967;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE D;
RX   MEDLINE=96283801; PubMed=8679691;
RA   Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA   Ogumad K.;
RT   "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT   types C and D organisms.";
RL   Biochim. Biophys. Acta 1307:123-126(1996).
DR   EMBL; D38442; BAA07477.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
```

```
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1285 AA;  147365 MW;  EEDE98E4ECA6C413 CRC64;

  Query Match             70.8%;  Score 1754;  DB 2;  Length 1285;
  Best Local Similarity   76.8%;  Pred. No. 7.4e-86;
  Matches  338;  Conservative   38;  Mismatches   56;  Indels    8;  Gaps    4;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
Db        851 SLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSS 910

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 145
Db        911 GDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSIGIIS 970

Qy        146 NFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIYINGKLIDTIKV 205
Db        971 NFLVFTLKQNENSEQDINFSYDISKNAAGYNKWFFVTITTNMMGNMMIYINGKLIDTIKV 1030

Qy        206 KELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQY 265
Db       1031 KELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSLQY 1090

Qy        266 TNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIR 325
Db       1091 TNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFNEGYKIIIKRIR 1150

Qy        326 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAIGLREQTKD-IND 383
Db       1151 GNTNDTRVRGENVLYFNTTIDNKQYSL-----GMYKPSRNLGTDLVPLGALDQPMDEIRK 1205

Qy        384 NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGD-WYRHNYLVPTVKQG 442
Db       1206 YGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLGDDYWFNHEYLIPVIKIE 1265

Qy        443 NYASLLESTSTHWGFVPVSE 462
Db       1266 HYASLLESTSTHWVFVPASE 1285


RESULT 4
Q841S3
ID   Q841S3      PRELIMINARY;      PRT;  1280 AA.
AC   Q841S3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2003;
RA   Lee M.S., Cheng M.C., Chen Y.P., Shiau J.R., Lin S.Y.;
RT   "An outbreak of botulism in black-faced spoonbill in Taiwan.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY251553; AAP06952.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
```

```
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1280 AA;  147642 MW;  34C1A2E573EE0AE7 CRC64;

  Query Match             32.9%;  Score 814.5;  DB 2;  Length 1280;
  Best Local Similarity   42.7%;  Pred. No. 1.4e-35;
  Matches  188;  Conservative   78;  Mismatches  145;  Indels   29;  Gaps   11;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
Db        855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
Db        915 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIKQNSGWKLCIR 971

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
Db        972 NGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSE 1031

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
Db       1032 RIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 1083

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
Db       1084 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPDRSKLYTGNPITIKS 1143

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHSTEDIYAIGLREQTKD 380
Db       1144 VSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSKNCVYA--LKSQSNL 1201

Qy        381 INDNI-IFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPT 438
Db       1202 GNYGIGIFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF------ENAYTP- 1252

Qy        439 VKQGNYASLLESTSTHWGFV 458
Db       1253 VAVTNYETKLLSTSSFWKFI 1272


RESULT 5
Q45849
ID   Q45849      PRELIMINARY;      PRT;  1280 AA.
AC   Q45849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum C.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=36828;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6813;
RX   MEDLINE=96156810; PubMed=8593068;
RA   Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA   Oguma K.;
RT   "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT   composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT   detection of this gene from Clostridium botulinum type C organisms.";
RL   Appl. Environ. Microbiol. 62:662-667(1996).
DR   EMBL; D49440; BAA08418.1; -.
DR   PIR; A43503; A43503.
```

```
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147815 MW;  74F813B228B8C989 CRC64;

  Query Match             32.8%;  Score 813.5;  DB 2;  Length 1280;
  Best Local Similarity   42.1%;  Pred. No. 1.5e-35;
  Matches  185;  Conservative   79;  Mismatches  148;  Indels   27;  Gaps   10;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :|||||||||||:|||||||||||:|| |||||||||||  |||||:| |:  |||| ||
Db        855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:| |||| : |
Db        915 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIKQNSGWKLCIR 971

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        972 NGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSE 1031

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
              ::::|  :   ||| | |::  |        :|  :|||| ||:|||  :||||::
Db       1032 RIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 1083

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
              ||:|||||| |::: |||::| :|::||:   |  :|     : :     |  | ||
Db       1084 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPDRSKLYTGNPITIKS 1143

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHSTEDIYAIGLREQTKD 380
              :      :|:  || : | |  |:  | :    :|:||       |   :||: |:    :
Db       1144 VSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSKNCVYALKLQSNLGN 1203

Qy        381 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 439
                 || |: : :   | |||| |    :|   :  |   :||          |  | |
Db       1204 YGIG-IFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF------ENAYTP-V 1253

Qy        440 KQGNYASLLESTSTHWGFV 458
                 || : | |||: | |:
Db       1254 AVTNYETKLLSTSSFWKFI 1272


RESULT 6
Q9LBS7
ID   Q9LBS7      PRELIMINARY;      PRT;  1280 AA.
AC   Q9LBS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-6814;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Kawabe T.,
RA   Murakami F., Nakatsuka M., Ohyama T.;
RT   "Organization of Gene Encoding Components of the Botulinum Progenitor
RT   Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT   Chimeric Sequence in the Gene Encoding Each Component.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037166; BAA89713.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147757 MW;  DE124FFB6F68450B CRC64;

  Query Match             32.8%;  Score 813.5;  DB 2;  Length 1280;
  Best Local Similarity   42.1%;  Pred. No. 1.5e-35;
  Matches  185;  Conservative   79;  Mismatches  148;  Indels   27;  Gaps   10;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :|||||||||||:|||||||||||:|| |||||||||||  |||||:| |:  |||| ||
Db        855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSS 914

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:| |||| : |
Db        915 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIKQNSGWKLCIR 971

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        972 NGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSE 1031

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
              ::::|  :   ||| | |::  |        :|  :|||| ||:|||  :||||::
Db       1032 RIEDLNEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 1083

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
              ||:|||||| |::: |||::| :|::||:   |  :|     : :     |  | ||
Db       1084 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPDRSKLYTGNPITIKS 1143

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYA---DNHSTEDIYAIGLREQTKD 380
              :      :|:  || : | |  |:  | :    :|:||       |   :||: |:    :
Db       1144 VSDKNPYSRILNGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSKNCVYALKLQSNLGN 1203

Qy        381 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 439
                 || |: : :   | |||| |    :|   :  |   :||          |  | |
Db       1204 YGIG-IFSIKNIVSQNKYCSQIFSSFM--KNTMLLADIYKPWRFSF------ENAYTP-V 1253

Qy        440 KQGNYASLLESTSTHWGFV 458
                 || : | |||: | |:
Db       1254 AVTNYETKLLSTSSFWKFI 1272


RESULT 7
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match             32.3%;  Score 801.5;  DB 12;  Length 1275;
  Best Local Similarity   41.7%;  Pred. No. 6.8e-35;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 85
              :|||||||||||:|||||||||||:|| ||||||||||    :|||| |:  |||| ||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 144
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        911 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 967

Qy        145 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 203
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        968 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 1027

Qy        204 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 263
              |:::|  :   ||| | |::  |        :|  :|||| ||:|||  :||||::
Db       1028 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 1079

Qy        264 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 323
              ||:|||||| |::: |||::| :|::||:   |  :|    : :    |  | ||
Db       1080 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKS 1139

Qy        324 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 380
              :      :|:  || :   |  |:: | :    :|:||      |   :||: |:   :
Db       1140 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 1199

Qy        381 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 439
                 || |: : :   | |||| |:|  ||   :  |   :||      ::: |    |
Db       1200 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 1251

Qy        440 KQGNYASLLESTSTHWGFV 458
                || : | |||: | |:
Db       1252 T--NYETKLLSTSSFWKFI 1268


RESULT 8
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match             22.9%;  Score 566.5;  DB 2;  Length 1252;
  Best Local Similarity   32.5%;  Pred. No. 2.5e-22;
  Matches  148;  Conservative   87;  Mismatches  151;  Indels   69;  Gaps   18;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    |::|  |  | :|::: : :  ||||||:: ::  |||   |    | |
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 892

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
              :  :| ::||: |:|:: |::||||||:||    || |:    ||||: :: ||||| :
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 952

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
              :  | :::||: |     | : |:|  :|    | ||| |||:||: :|: |:|||| ||
Db        953 SLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLI 1012

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |   :  |  |: |  | |:|    |  |              || | || |||| :|  |:
Db       1013 DKKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1063

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDF 313
              |:    |::||:||| | |:||||::|:    :::||    :|   : | |       |
Db       1064 NNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRL 1121

Qy        314 NEGYKIIIKRI-RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 372
              | |: |:|:   :|||  ||  | :| :  : :| : |       :|||  :|
Db       1122 YSGIKVKIQRVNNSSTNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTN----- 1170

Qy        373 GLREQTKDIND--NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LG 426
              :|:|  |:   |  |:  ||:     :  ||:| || |       ||   |:    :
Db       1171 --KEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVA 1221

Qy        427 GDWYRHNYLVPTVKQGNYASLLESTSTH---WGFV 458
                ||              |  : ::|:::   | |:
Db       1222 STWY-------------YTHMRDNTNSNGFFWNFI 1243


RESULT 9
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match             22.9%;  Score 566.5;  DB 2;  Length 1255;
  Best Local Similarity   32.5%;  Pred. No. 2.5e-22;
  Matches  148;  Conservative   87;  Mismatches  151;  Indels   69;  Gaps   18;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
Db        838 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQF--GIYN 895

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
Db        896 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKV 955

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
Db        956 SLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLI 1015

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
Db       1016 DKKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1066

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNI----DYLNRYMYANSRQIVFNTRRN---NNDF 313
Db       1067 NNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFINR--RTDSTLSINNIRSTILLANRL 1124

Qy        314 NEGYKIIIKRI-RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 372
Db       1125 YSGIKVKIQRVNNSSTNDNLVRKNDQVYINF-VASKTHLL-----PLYADTATTN----- 1173

Qy        373 GLREQTKDIND--NIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LG 426
Db       1174 --KEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVA 1224

Qy        427 GDWYRHNYLVPTVKQGNYASLLESTSTH---WGFV 458
Db       1225 STWY-------------YTHMRDNTNSNGFFWNFI 1246


RESULT 10
Q933K0
```

```
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             22.8%;  Score 565;  DB 2;  Length 1291;
  Best Local Similarity   33.3%;  Pred. No. 3.1e-22;
  Matches  161;  Conservative   79;  Mismatches  153;  Indels   90;  Gaps   20;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140
Db        906 --NSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
Db        964 ISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKL 1022

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN----MWIRDFYIFAKELDGKD 255
Db       1023 ESHIDIRDIREVIANDEIIFKLD-----------GNIDRTQFIWMKYFSIFNTELSQSN 1070

Qy        256 INILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN 311
Db       1071 IEEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQN 1130

Qy        312 -------DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFDMTINNKAYNLFMKNETMYA 361
Db       1131 SKYINYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYM------- 1182

Qy        362 DNHSTEDIYAIGLREQTK----DINDNIIF----QIQPMNNTYYYASQ-IFKSNFNGENI 412
Db       1183 --------YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKK--DEEST 1232

Qy        413 SGICSIGTYRFRLGG-------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVP 459
Db       1233 DEIGLIGIHRFYESGIVFKEYKDYFCISKWY-----LKEVKRKPYNSKL---GCNWQFIP 1284

Qy        460 VSE 462
```

```
Db       1285 KDE 1287


RESULT 11
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match             22.8%;  Score 565;  DB 2;  Length 1291;
  Best Local Similarity   32.4%;  Pred. No. 3.1e-22;
  Matches  157;  Conservative   86;  Mismatches  157;  Indels   84;  Gaps   20;

Qy         23 ASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKL 82
              ::: :|  | |:| : | :: ||:|: | | |:| ||| |:|     |:||      |||
Db        844 SNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKL 901

Qy         83 GSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNN 136
              || :   |: ||||:||::|||:  ||:|||||| |:    : |     ||||: :|||
Db        902 TSSAD--SKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNN 959

Qy        137 SGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYI 195
              ||| | |  | :::||       :|: | |:|  :   | |:|||||:||| : | ||||
Db        960 SGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYI 1018

Qy        196 NGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGK 254
              || |   : :|::  :  :  |||:::          | |   :|:: | ||  :|:
Db       1019 NGTLESNMDIKDIGEVIVNGEITFKLD---------GDVDRTQFIWMKYFSIFNTQLNQS 1069

Qy        255 DINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNN 310
              :|  ::    |:  :||:||| | ||||||| |    | |:     ::  :|:  :: |
Db       1070 NIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQ 1129

Qy        311 N-------DFNEGYKIIIKRIRG--NTNDTRVRGGDILYFDMTINNK-----AYNLFMKN 356
              |       :   | | ||:|     : ||  ||  | :: |: ::::     ||  | :
Db       1130 NSNYINYRNLYIGEKFIIRRESNSQSINDDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQ 1189

Qy        357 E-----TMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGEN 411
              |     :: :|  | |    | ::|          :  ||     |    :||   : |:
Db       1190 EEKLFLSIISD--SNEFYKTIEIKE----------YDEQPS----YSCQLLFKK--DEES 1231

Qy        412 ISGICSIGTYRFRLGG-------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFV 458
                  | || :||   |               ||     :  ||:  | | |     :| |:
Db       1232 TDDIGLIGIHRFYESGVLRKKYKDYFCISKWY-----LKEVKRKPYKSNL---GCNWQFI 1283

Qy        459 PVSE 462
              |  |
Db       1284 PKDE 1287


RESULT 12
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match             22.5%;  Score 559;  DB 2;  Length 1291;
  Best Local Similarity   33.1%;  Pred. No. 6.5e-22;
  Matches  160;  Conservative   79;  Mismatches  154;  Indels   90;  Gaps   20;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :| :| |:| ::| :: ||:|: | | |:| ||| |:|     |:||      ||| ||
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140
                 |: | ||:||::|||:  ||:|||||| |:    : |     ||||: :||||||
Db        906 --NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
              | |  | :::||       :|: | | |  :   | |:|||||:|||   | ||||||||
Db        964 ISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKL 1022

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN----MWIRDFYIFAKELDGKD 255
              | ::::  :  :  | |:::            ||:    :|:: | ||  ||   :
Db       1023 ESHIDIRDIREVIANDEIIFKLD-----------GNIDRTQFIWMKYFSIFNTELSQSN 1070
```

```
Qy        256 INILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN 311
Db       1071 IEEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQN 1130

Qy        312 -------DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFDMTINNKAYNLFMKNETMYA 361
Db       1131 SKYINYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYM------- 1182

Qy        362 DNHSTEDIYAIGLREQTK----DINDNIIF----QIQPMNNTYYYASQ-IFKSNFNGENI 412
Db       1183 --------YKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKK--DEEST 1232

Qy        413 SGICSIGTYRFRLGG-------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVP 459
Db       1233 DEIGLIGIHRFYESGIVFKEYKDYFCISKWY-----LKEVKRKPYNSKL---GCNWQFIP 1284

Qy        460 VSE 462
Db       1285 KDE 1287
```

RESULT 13
Q8GR96
ID Q8GR96 PRELIMINARY; PRT; 1291 AA.
AC Q8GR96;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin.
GN BONTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA Nakamura S., Mukamoto M., Kozaki S.;
RT "Clostridium botulinum type B neurotoxin associated with infant
RT botulism.";
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB084152; BAC22064.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150574 MW; 0227CAEF4F58504D CRC64;

Query Match 22.5%; Score 559; DB 2; Length 1291;
Best Local Similarity 33.2%; Pred. No. 6.5e-22;
Matches 159; Conservative 76; Mismatches 162; Indels 82; Gaps 21;

```
Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
Db        848 ILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELND--KNQFKLTSS- 904

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140
Db        905 -TNSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWK 963

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199
Db        964 ISIRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNN-SDNAKIYINGKL 1022

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN-MWIRDFYIFAKELDGKDINI 258
Db       1023 ESNIDIKDIGEVIANGEIIFKLD---------GDIDRTQFIWMKYFSIFNTELSQSNIKE 1073

Qy        259 LFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMY----ANSRQIVFNTRRNNN--- 311
Db       1074 IYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNY 1133

Qy        312 ----DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFDMTINNK-----AYNLFMKNE-- 357
Db       1134 INYRNLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKK 1192

Qy        358 TMYADNHSTEDIY-AIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGIC 416
Db       1193 LFLANIYDSNEFYKTIQIKE----------YDEQPT----YSCQLLFKK--DEESTDEIG 1236

Qy        417 SIGTYRFRLGG-------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE 462
Db       1237 LIGIHRFYESGIVLKDYKNYFCISKWY-----LKEVKRKPYNPNL---GCNWQFIPKDE 1287
```

RESULT 14
Q9K395
ID Q9K395 PRELIMINARY; PRT; 1251 AA.
AC Q9K395;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 095;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 155;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA Kato H., Nakamura S., Karasawa T.;
RT "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1899;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1897;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1898;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.

```
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             22.5%;  Score 558;  DB 2;  Length 1251;
  Best Local Similarity   31.8%;  Pred. No. 7.1e-22;
  Matches  144;  Conservative   90;  Mismatches  153;  Indels   66;  Gaps   16;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
              :|    |::|  |  | :|::: : :  ||||||:: ::  |:: : |     | : :|
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFNS- 893

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRI----NKWVSNLPGYTIIDSVK-NNSGWSI 141
                 :| ::||: |:|:: |::|||||:||    || |:    ||||: :: ||||| :
Db        894 -KPSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNINNEYTIINCMRDNNSGWKV 952

Qy        142 GIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLI 200
               :  | :::||: |    | : | |  :|    | ||| |||:||: :|: |:|||| ||
Db        953 SLNHNEIIWTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGHLI 1012

Qy        201 DTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILF 260
              |  :  |  |: |  | |:|     |  |            || | || ||||  :|  |:
Db       1013 DQKSILNLGNIHVSDNILFKIVNCSYTRYI---------GIRYFNIFDKELDETEIQTLY 1063

Qy        261 NSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYM--YANSRQIVFNTRRN---NNDFNE 315
              ::   ||::||:||| | |:| ||::|:   | ::    :|   : | |     |
Db       1064 SNEPNTNILKDFWGNYLLYDKGYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYS 1123

Qy        316 GYKIIIKRIR-GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGL 374
              | |: |:|:   :|||  ||  | :| :   |: :|:|       ||| ::|:
Db       1124 GIKVKIQRVNDSSTNDRFVRKNDQVYINYISNSSSYSL-------YADTNTTD------- 1169

Qy        375 REQT--KDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFR----LGGD 428
              :|:|    : |   |:  ||:     :  ||:| || |     ||   |:    :
Db       1170 KEKTIKSSSSGNRFNQVVVMNSVGNNCTMNFKNN-NGNN------IGLLGFKADTVVAST 1222

Qy        429 WYRHNYLVPTVKQGNYASLLESTSTH---WGFV 458
              ||              |  : : |:::   | |:
Db       1223 WY-------------YTHMRDHTNSNGCFWNFI 1242


RESULT 15
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             22.5%;  Score 557;  DB 2;  Length 1291;
  Best Local Similarity   33.1%;  Pred. No. 8.3e-22;
  Matches  160;  Conservative   77;  Mismatches  156;  Indels   90;  Gaps   20;

Qy         27 LLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSG 86
```

```
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         87 EDRGKVIVTQNENIVYNSMYESFSISFWIRINKW----VSNL--PGYTIIDSVKNNSGWS 140

Db        906 --NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        141 IGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKL 199

Db        964 ISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKL 1022

Qy        200 IDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN----MWIRDFYIFAKELDGKD 255

Db       1023 ESHIDIRDIREVIANDEIIFKLD-----------GNIDRTQFIWMKYFSIFNTELSQSN 1070

Qy        256 INILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANS----RQIVFNTRRNNN 311

Db       1071 IEEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQN 1130

Qy        312 -------DFNEGYKIIIKRIRGNT---NDTRVRGGDILYFDMTINNKAYNLFMKNETMYA 361

Db       1131 SKYINYRDLYIGEKFIIRR-KSNSQSINDDIVRKEDYIYLDFFNLNQEWRVY-------- 1181

Qy        362 DNHSTEDIYAIGLREQTK----DINDNIIF----QIQPMNNTYYYASQ-IFKSNFNGENI 412

Db       1182 -------IYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKK--DEEST 1232

Qy        413 SGICSIGTYRFRLGG------------DWYRHNYLVPTVKQGNYASLLESTSTHWGFVP 459

Db       1233 DEIGLIGIHRFYESGIVFKEYKDYFCISKWY-----LKEVKRKPYNSKL---GCNWQFIP 1284

Qy        460 VSE 462

Db       1285 KDE 1287
```

Search completed: September 10, 2004, 03:41:44
Job time : 23.1388 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 26.4249 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 451
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	350	77.6	451	2	AAW68398	Aaw68398 Clostridi
2	270	59.9	451	4	AAB04093	Aab04093 Botulism
3	37	8.2	462	2	AAW68397	Aaw68397 Clostridi
4	27	6.0	451	2	AAW68395	Aaw68395 Clostridi
5	27	6.0	452	2	AAW68396	Aaw68396 Clostridi
6	26	5.8	448	2	AAW68399	Aaw68399 Clostridi
7	26	5.8	462	2	AAR95009	Aar95009 Type A ne
8	26	5.8	462	2	AAW68390	Aaw68390 Clostridi
9	26	5.8	472	2	AAW68393	Aaw68393 Clostridi
10	26	5.8	472	2	AAW68394	Aaw68394 Clostridi
11	26	5.8	473	2	AAW68400	Aaw68400 Clostridi
12	23	5.1	393	4	AAG68149	Aag68149 Streptomy
13	23	5.1	393	4	AAB74844	Aab74844 S. griseu
14	23	5.1	423	2	AAW68098	Aaw68098 Chlamydom
15	22	4.9	22	4	AAB67674	Aab67674 Amino aci
16	22	4.9	22	5	AAM49518	Aam49518 His-tag p
17	22	4.9	26	7	ADD31745	Add31745 Purificat
18	22	4.9	53	2	AAW36986	Aaw36986 HFKT pept
19	22	4.9	54	2	AAW36990	Aaw36990 HFKT pept
20	22	4.9	91	2	AAY09127	Aay09127 Recombina
21	22	4.9	93	2	AAY09128	Aay09128 Recombina
22	22	4.9	195	3	AAB11377	Aab11377 B. stearo
23	22	4.9	269	5	AAM49520	Aam49520 B. mori 1
24	22	4.9	269	5	AAM49541	Aam49541 B. mori T
25	22	4.9	323	2	AAW71185	Aaw71185 FP505 pro
26	22	4.9	332	4	AAE02191	Aae02191 Staphyloc
27	22	4.9	364	3	AAB13011	Aab13011 Q-tagged
28	22	4.9	567	4	AAU00220	Aau00220 Bad-DTTR
29	21	4.7	21	2	AAW23653	Aaw23653 Recombina
30	21	4.7	21	2	AAW23650	Aaw23650 Recombina
31	21	4.7	21	2	AAW23647	Aaw23647 Recombina
32	21	4.7	194	5	AAM51139	Aam51139 HIV TAT t
33	20	4.4	411	4	AAU00219	Aau00219 Bcl-Xl-DT
34	17	3.8	24	2	AAW19779	Aaw19779 Polyhisti
35	17	3.8	24	2	AAY06647	Aay06647 Polyhisti
36	17	3.8	24	2	AAY27015	Aay27015 Met Gly H
37	17	3.8	24	2	AAY09370	Aay09370 12 amino
38	17	3.8	24	3	AAY81959	Aay81959 Enterokin
39	17	3.8	24	4	AAB47057	Aab47057 Peptide f
40	17	3.8	24	6	ABG73858	Abg73858 Plasmid p
41	17	3.8	24	6	ABU10203	Abu10203 Peptide f
42	17	3.8	128	6	ABB99729	Abb99729 Amino aci
43	17	3.8	135	5	ABJ15133	Abj15133 Synthetic
44	17	3.8	135	5	ABJ15135	Abj15135 HisNeubla
45	15	3.3	20	5	ABB76190	Abb76190 Modified

ALIGNMENTS

RESULT 1
AAW68398
ID AAW68398 standard; protein; 451 AA.
XX
AC AAW68398;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type D toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW botulism; BotD.
XX
OS Clostridium botulinum; serotype D.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30591.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 47; Page 352-353; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC BotD soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which

```
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             77.6%;  Score 350;  DB 2;  Length 451;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  450;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNLLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 2
AAB04093
ID   AAB04093 standard; protein; 451 AA.
XX
AC   AAB04093;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype D).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54487.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 6b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 451 AA;

  Query Match             59.9%;  Score 270;  DB 4;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-269;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLV 296
              |||||||||||||||||||||||||||||
Db        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLV 296


RESULT 3
AAW68397
ID   AAW68397 standard; protein; 462 AA.
XX
AC   AAW68397;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type C1 toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW   botulism; BotC.
XX
OS   Clostridium botulinum; serotype C1 strain Stockholm.
```

OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30588.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 45; Page 339-341; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC express BotC soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 462 AA;

Query Match 8.2%; Score 37; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 5.3e-29;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFN 37
||||||||||||||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFN 37

RESULT 4
AAW68395
ID AAW68395 standard; protein; 451 AA.
XX
AC AAW68395;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain Belgua.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30584.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 41; Page 324-325; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC express BotE soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 451 AA;

Query Match 6.0%; Score 27; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.1e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27
|||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27

RESULT 5
AAW68396
ID AAW68396 standard; protein; 452 AA.
XX
AC AAW68396;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain NCTC 11219.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX

PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30585.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 41; Page 327-329; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC express BotE soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins,
CC especially type B and type E toxins. Methods are provided which allow for
CC the isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The recombinant
CC toxins are used as immunogens for the production of vaccines and
CC antitoxins that are useful in the treatment of humans and animals at risk
CC of intoxication with clostridial toxin
XX
SQ Sequence 452 AA;

Query Match 6.0%; Score 27; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.1e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMAL 27

RESULT 6
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30593.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC express BotC soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX
SQ Sequence 448 AA;

Query Match 5.8%; Score 26; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.2e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 7
AAR95009
ID AAR95009 standard; protein; 462 AA.
XX
AC AAR95009;
XX
DT 07-JUL-1996 (first entry)
XX
DE Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /label= Polyhistidine_tag
FT Protein 22. .462
FT /label= C-fragment
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29246.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.

PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 7; Page 340-342; 434pp; English.
XX
CC pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC fragment C (see also AAR95008) of the Clostridium botulinum type A
CC neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC soluble protein and was purified by metal chelate affinity chromatography
CC to obtain a product free of endotoxin contamination that may be useful as
CC an immunogen in vaccine compsns
XX
SQ Sequence 462 AA;

Query Match 5.8%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.2e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 8
AAW68390
ID AAW68390 standard; protein; 462 AA.
XX
AC AAW68390;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism.
XX
OS Clostridium botulinum; serotype A.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal histidine tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30572, AAV30575.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 24; Page 265-267; 428pp; English.
XX
CC This is the amino acid sequence of a histidine-tagged fragment C
CC polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on a poly-histidine binding
CC affinity resin. The invention relates to recombinant proteins derived
CC from C. botulinum toxins. Methods are provided which allow the isolation
CC of soluble recombinant proteins that are free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant proteins are
CC E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are useful
CC in the treatment of humans and animals at risk of intoxication with
CC clostridial toxin
XX
SQ Sequence 462 AA;

Query Match 5.8%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.2e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 9
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

```
  Query Match             5.8%;  Score 26;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 10
AAW68394
ID   AAW68394 standard; protein; 472 AA.
XX
AC   AAW68394;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30581.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 303-305; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
CC   animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 472 AA;

  Query Match             5.8%;  Score 26;  DB 2;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 11
AAW68400
ID   AAW68400 standard; protein; 473 AA.
XX
AC   AAW68400;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW   botulism; BotG.
XX
OS   Clostridium botulinum; serotype G strain 113/30.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30596.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC   sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC   express BotG soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match             5.8%;  Score 26;  DB 2;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 12
AAG68149
ID   AAG68149 standard; protein; 393 AA.
XX
AC   AAG68149;
```

XX
DT 24-JAN-2002 (first entry)
XX
DE Streptomyces griseus RppA protein with a His tag SEQ ID NO:5.
XX
KW Streptomyces griseus; RppA; pyran-2-one; drug; agricultural.
XX
OS Streptomyces griseus.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /label= His_tag
FT Protein 22. .393
FT /label= RppA_protein
XX
PN JP2001238677-A.
XX
PD 04-SEP-2001.
XX
PF 01-MAR-2000; 2000JP-00055782.
XX
PR 01-MAR-2000; 2000JP-00055782.
XX
PA (MITR) MITSUBISHI RAYON CO LTD.
XX
DR WPI; 2001-613891/71.
XX
PT Preparation of pyran-2-one derivative, used as intermediate and as
PT additive for thermosensible transfer paper, comprises contacting carbonyl
PT compound with microbe containing polyketide synthase gene.
XX
PS Example 1; Page 10-11; 11pp; Japanese.
XX
CC The present invention describes the preparation of pyran-2-one
CC derivatives. The preparation method comprises contacting carbonyl
CC compounds with microbes containing a polyketide synthase gene. Pyran-2-
CC one derivatives can be used as synthetic intermediates for drugs and
CC agricultural chemicals. They can also be used as additives for
CC thermosensible transfer paper. The present sequence represents the
CC Streptomyces griseus RppA protein with an N-terminal His tag, which is
CC used in an example from the present invention
XX
SQ Sequence 393 AA;

Query Match 5.1%; Score 23; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.3e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

RESULT 13
AAB74844
ID AAB74844 standard; protein; 393 AA.
XX
AC AAB74844;
XX
DT 18-JUN-2001 (first entry)
XX
DE S. griseus actinomycetes polyketide synthase with a histidine tag.
XX
KW Streptomyces griseus; actinomycete; polyketide synthase; rppA; drug;
KW 1,3,6,8-tetrahydroxynaphthalene; melanine; agricultural chemical;
KW malonyl CoA.
XX
OS Streptomyces griseus.
OS Synthetic.
XX
PN JP2000342269-A.
XX
PD 12-DEC-2000.
XX
PF 02-JUN-1999; 99JP-00155526.
XX
PR 02-JUN-1999; 99JP-00155526.
XX
PA (MITR) MITSUBISHI RAYON CO LTD.
XX
DR WPI; 2001-303158/32.
XX
PT Preparation of 1,3,6,8-tetrahydroxynaphthalene for use as a synthetic
PT material for melanine and as an intermediate for drugs and agricultural
PT chemicals, comprises using a microbe and an enzyme.
XX
PS Example 1; Page 9-10; 10pp; Japanese.
XX
CC The present invention describes a method for preparing 1,3,6,8-
CC tetrahydroxynaphthalene comprising using a microbe cell containing an
CC rppA gene encoding actinomycetes polyketide synthase, to convert malonyl
CC CoA as the substrate or a substance which can produce malonyl CoA in the
CC microbe cell to 1,3,6,8-tetrahydroxynaphthalene and recovering it.
CC 1,3,6,8-tetrahydroxynaphthalene is used as a synthetic material for
CC melanine and as an intermediate for drugs and agricultural chemicals. The
CC present sequence represents the actinomycetes polyketide synthase protein
CC isolated from Streptomyces griseus comprising an histidine tag, which is
CC used in an example from the present invention
XX
SQ Sequence 393 AA;

Query Match 5.1%; Score 23; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.3e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

RESULT 14
AAW68098
ID AAW68098 standard; protein; 423 AA.
XX
AC AAW68098;
XX
DT 27-OCT-1998 (first entry)
XX
DE Chlamydomonas reinhardtii histidine tagged RB47 binding protein.
XX
KW RB47; activator; mRNA binding protein; psbA; chloroplast; recombinant;
KW regulated production; protein synthesis; plant propagation.
XX
OS Chlamydomonas reinhardtii.
OS Synthetic.
XX
PN WO9831823-A1.
XX
PD 23-JUL-1998.
XX
PF 16-JAN-1998; 98WO-US000840.
XX
PR 17-JAN-1997; 97US-0035955P.
PR 12-DEC-1997; 97US-0069400P.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Mayfield S;
XX
DR WPI; 1998-427562/36.
DR N-PSDB; AAV41263.
XX
PT New isolated RB47 and RB60 genes - used for, e.g. regulated expression of
PT desired recombinant molecule in prokaryote and eukaryote, particularly
PT plant cells.
XX

PS Claim 10; Fig 5; 109pp; English.
XX
CC The sequence is that of a histidine tagged RB47 mRNA binding protein. It
CC can be used in the regulated production of a desired recombinant
CC molecule. It can be used as a cell culture system, an in vitro expression
CC system or a whole tissue, preferably a plant and allows for a high level
CC of protein synthesis in plant chloroplasts and in the cytoplasm of both
CC prokaryotic and eukaryotic cells. It can also be used to screen for
CC agonists or antagonists of the binding of RB47 to the RB47 binding site.
CC Antagonists of the binding are useful in the prevention of plant
CC propagation. They can also be used for screening for agonists or
CC antagonists of RB60. Such agonists or antagonists would be useful in
CC general to modify expression of RB60 as a way to regulate cellular
CC processes in a redox manner
XX
SQ Sequence 423 AA;

Query Match 5.1%; Score 23; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

RESULT 15
AAB67674
ID AAB67674 standard; peptide; 22 AA.
XX
AC AAB67674;
XX
DT 11-JUN-2001 (first entry)
XX
DE Amino acid sequence of His tag and factor Xa protease cleavage site.
XX
KW HIV; tat protein; effector protein; transduction domain; factor Xa.
XX
OS Synthetic.
XX
PN WO200119393-A1.
XX
PD 22-MAR-2001.
XX
PF 13-SEP-2000; 2000WO-US024977.
XX
PR 13-SEP-1999; 99US-0153507P.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Collmer A, Beer SV;
XX
DR WPI; 2001-257850/26.
DR N-PSDB; AAF55675.
XX
PT Delivering effector proteins into target cell for use in protein therapy,
PT involves introducing effector protein fused to protein transduction
PT domain of human immunodeficiency virus TAT protein, into target cells.
XX
PS Example 1; Page 20; 43pp; English.
XX
CC The present sequence represents a His tag and factor Xa protease cleavage
CC site, which is used in the method of the invention. The specification
CC describes a method for delivering effector proteins into a target cell.
CC The method comprises introducing an effector protein fused to a protein
CC transduction domain of a human immunodeficiency virus (HIV) tat protein
CC into the target cell. The method is used for delivering a heterologous
CC effector protein such as an effector protein produced by bacterial plant
CC pathogen, animal pathogen or a rhizosphere bacteria, or a protein
CC secreted and/or delivered into eukaryotic cells by a Type III secretion
CC system or a hypersensitive response elicitor, an avirulence protein, a
CC hypersensitive response and pathogenicity-dependent outer protein, a
CC virulence protein or a pathogenicity protein, into an eukaryotic cell.
CC The method is useful for delivering effector proteins for use in
CC pharmaceutical, insecticidal, fungicide, herbicide and other applications
XX
SQ Sequence 22 AA;

Query Match 4.9%; Score 22; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 8.6e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHM 22
||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHM 22

Search completed: September 10, 2004, 05:06:01
Job time : 27.4249 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 7.39438 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 451
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	5.8	462	1	US-08-480-604A-26	Sequence 26, Appl
2	26	5.8	462	2	US-08-405-496A-26	Sequence 26, Appl
3	26	5.8	462	3	US-08-915-136-26	Sequence 26, Appl
4	26	5.8	462	4	US-09-084-517-26	Sequence 26, Appl
5	23	5.1	23	1	US-08-480-604A-24	Sequence 24, Appl
6	23	5.1	23	2	US-08-405-496A-24	Sequence 24, Appl
7	23	5.1	23	3	US-08-915-136-24	Sequence 24, Appl
8	23	5.1	23	4	US-09-084-517-24	Sequence 24, Appl
9	22	4.9	22	4	US-09-660-742-3	Sequence 3, Appli
10	22	4.9	53	2	US-08-651-818A-19	Sequence 19, Appl
11	22	4.9	53	3	US-09-184-826-19	Sequence 19, Appl
12	22	4.9	54	2	US-08-651-818A-23	Sequence 23, Appl
13	22	4.9	54	3	US-09-184-826-23	Sequence 23, Appl
14	22	4.9	323	1	US-08-591-989-5	Sequence 5, Appli
15	21	4.7	21	2	US-08-651-818A-21	Sequence 21, Appl
16	21	4.7	21	3	US-09-184-826-21	Sequence 21, Appl
17	17	3.8	24	3	US-08-584-031-8	Sequence 8, Appli
18	17	3.8	24	3	US-08-780-496-8	Sequence 8, Appli
19	17	3.8	24	4	US-08-928-069-9	Sequence 9, Appli
20	15	3.3	20	4	US-09-186-489-6	Sequence 6, Appli
21	15	3.3	20	4	US-10-043-665B-6	Sequence 6, Appli
22	12	2.7	16	5	PCT-US94-04361-56	Sequence 56, Appl
23	12	2.7	17	2	US-08-600-783-13	Sequence 13, Appl
24	11	2.4	20	4	US-09-674-677-34	Sequence 34, Appl
25	11	2.4	24	4	US-09-292-858B-27	Sequence 27, Appl
26	11	2.4	254	3	US-09-167-434-7	Sequence 7, Appli
27	11	2.4	254	3	US-08-853-755-7	Sequence 7, Appli
28	11	2.4	302	2	US-08-203-532F-4	Sequence 4, Appli
29	11	2.4	302	3	US-08-950-860-16	Sequence 16, Appl
30	11	2.4	302	3	US-09-078-465-4	Sequence 4, Appli
31	11	2.4	302	5	PCT-US95-01882A-4	Sequence 4, Appli
32	11	2.4	303	2	US-08-203-532F-2	Sequence 2, Appli
33	11	2.4	303	3	US-09-078-465-2	Sequence 2, Appli
34	11	2.4	303	5	PCT-US95-01882A-2	Sequence 2, Appli
35	11	2.4	363	4	US-09-328-352-4930	Sequence 4930, Ap
36	11	2.4	414	5	PCT-US92-06840-2	Sequence 2, Appli
37	10	2.2	25	4	US-09-721-154-14	Sequence 14, Appl
38	10	2.2	335	3	US-09-095-117-2	Sequence 2, Appli
39	10	2.2	335	4	US-09-850-260-2	Sequence 2, Appli
40	10	2.2	449	2	US-08-927-394-2	Sequence 2, Appli
41	10	2.2	480	1	US-07-882-292-2	Sequence 2, Appli
42	10	2.2	480	2	US-08-331-644-2	Sequence 2, Appli
43	10	2.2	480	5	PCT-US93-04102-2	Sequence 2, Appli
44	10	2.2	633	3	US-08-557-006C-43	Sequence 43, Appl
45	10	2.2	763	2	US-08-677-862-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             5.8%;  Score 26;  DB 1;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match             5.8%;  Score 26;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             5.8%;  Score 26;  DB 3;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4e-17;
```

```
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             5.8%;  Score 26;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 5
US-08-480-604A-24
; Sequence 24, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-24

  Query Match             5.1%;  Score 23;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 6
US-08-405-496A-24
```

```
; Sequence 24, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-24

  Query Match             5.1%;  Score 23;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 7
US-08-915-136-24
; Sequence 24, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-24

  Query Match             5.1%;  Score 23;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 8
US-09-084-517-24
; Sequence 24, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
```

```
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,517
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CARROLL, PETER G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  OPHD-01610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  23 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-084-517-24

  Query Match             5.1%;  Score 23;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 9
US-09-660-742-3
; Sequence 3, Application US/09660742
; Patent No. 6485977
; GENERAL INFORMATION:
;  APPLICANT:  Collmer, Alan
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  RECOMBINANT CONSTRUCTS AND TECHNIQUES FOR DELIVERING TO
;  TITLE OF INVENTION:  EUCARYOTIC CELLS BACTERIAL PROTEINS THAT ARE SECRETED
;  TITLE OF INVENTION:  VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE:  19603/4021
;  CURRENT APPLICATION NUMBER: US/09/660,742
;  CURRENT FILING DATE:  2000-09-13
;  PRIOR APPLICATION NUMBER:  60/153,507
;  PRIOR FILING DATE:  1999-09-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 22
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: · Polypeptide
US-09-660-742-3

  Query Match             4.9%;  Score 22;  DB 4;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 2e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 10
US-08-651-818A-19
; Sequence 19, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-19

  Query Match             4.9%;  Score 22;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 11
US-09-184-826-19
; Sequence 19, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
```

```
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-19

  Query Match             4.9%;  Score 22;  DB 3;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 12
US-08-651-818A-23
; Sequence 23, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-23

  Query Match             4.9%;  Score 22;  DB 2;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 13
US-09-184-826-23
; Sequence 23, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-23

  Query Match             4.9%;  Score 22;  DB 3;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 14
US-08-591-989-5
```

```
; Sequence 5, Application US/08591989
; Patent No. 5795721
;  GENERAL INFORMATION:
;    APPLICANT:  Ross S. Rabin, Sumedha Jayasena
;    APPLICANT:  and Larry Gold
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC
;    TITLE OF INVENTION:  ACID LIGANDS OF ICP4
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 East Prentice Avenue, Suite #200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.40 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX 49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-591-989-5

  Query Match             4.9%;  Score 22;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db         17 MGHHHHHHHHHHSSGHIEGRHM 38


RESULT 15
US-08-651-818A-21
; Sequence 21, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-21

  Query Match             4.7%;  Score 21;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-13;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRH 21
              |||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRH 21


Search completed: September 10, 2004, 05:17:09
Job time : 7.39438 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 22.8137 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 451
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	451	100.0	451	12	US-10-272-898-68	Sequence 68, Appl
2	451	100.0	451	15	US-10-354-774-68	Sequence 68, Appl
3	451	100.0	451	15	US-10-271-012-68	Sequence 68, Appl
4	451	100.0	451	16	US-10-729-122-68	Sequence 68, Appl
5	451	100.0	451	16	US-10-729-039-68	Sequence 68, Appl
6	425	94.2	1276	12	US-10-272-898-66	Sequence 66, Appl
7	425	94.2	1276	15	US-10-354-774-66	Sequence 66, Appl
8	425	94.2	1276	15	US-10-271-012-66	Sequence 66, Appl
9	425	94.2	1276	15	US-10-452-024-164	Sequence 164, App
10	425	94.2	1276	16	US-10-729-122-66	Sequence 66, Appl
11	425	94.2	1276	16	US-10-729-039-66	Sequence 66, Appl
12	324	71.8	1276	15	US-10-452-024-4	Sequence 4, Appli
13	324	71.8	1276	15	US-10-452-024-168	Sequence 168, App
14	324	71.8	1276	15	US-10-205-516-8	Sequence 8, Appli
15	324	71.8	1286	15	US-10-205-516-22	Sequence 22, Appl
16	323	71.6	1275	15	US-10-452-024-167	Sequence 167, App
17	270	59.9	451	10	US-09-910-186A-12	Sequence 12, Appl
18	80	17.7	1280	15	US-10-452-024-139	Sequence 139, App
19	80	17.7	1280	15	US-10-452-024-140	Sequence 140, App
20	37	8.2	462	12	US-10-272-898-62	Sequence 62, Appl
21	37	8.2	462	15	US-10-354-774-62	Sequence 62, Appl
22	37	8.2	462	15	US-10-271-012-62	Sequence 62, Appl
23	37	8.2	462	16	US-10-729-122-62	Sequence 62, Appl
24	37	8.2	462	16	US-10-729-039-62	Sequence 62, Appl
25	27	6.0	451	12	US-10-272-898-54	Sequence 54, Appl
26	27	6.0	451	15	US-10-354-774-54	Sequence 54, Appl
27	27	6.0	451	15	US-10-271-012-54	Sequence 54, Appl
28	27	6.0	451	16	US-10-729-122-54	Sequence 54, Appl
29	27	6.0	451	16	US-10-729-039-54	Sequence 54, Appl
30	27	6.0	452	12	US-10-272-898-56	Sequence 56, Appl
31	27	6.0	452	15	US-10-354-774-56	Sequence 56, Appl
32	27	6.0	452	15	US-10-271-012-56	Sequence 56, Appl
33	27	6.0	452	16	US-10-729-122-56	Sequence 56, Appl
34	27	6.0	452	16	US-10-729-039-56	Sequence 56, Appl
35	26	5.8	448	12	US-10-272-898-73	Sequence 73, Appl
36	26	5.8	448	15	US-10-354-774-73	Sequence 73, Appl
37	26	5.8	448	15	US-10-271-012-73	Sequence 73, Appl
38	26	5.8	448	16	US-10-729-122-73	Sequence 73, Appl
39	26	5.8	448	16	US-10-729-039-73	Sequence 73, Appl
40	26	5.8	462	12	US-10-272-898-26	Sequence 26, Appl
41	26	5.8	462	15	US-10-354-774-26	Sequence 26, Appl
42	26	5.8	462	15	US-10-271-012-26	Sequence 26, Appl
43	26	5.8	462	16	US-10-729-122-26	Sequence 26, Appl
44	26	5.8	462	16	US-10-729-039-26	Sequence 26, Appl
45	26	5.8	472	12	US-10-272-898-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-68
; Sequence 68, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-272-898-68

  Query Match             100.0%;  Score 451;  DB 12;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 2
US-10-354-774-68
; Sequence 68, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 68:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-354-774-68

  Query Match             100.0%;  Score 451;  DB 15;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 3
US-10-271-012-68
; Sequence 68, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/271,012
;              FILING DATE: 15-Oct-2002
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US/08/704,159
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Ingolia, Diane E.
;              REGISTRATION NUMBER: 40,027
;              REFERENCE/DOCKET NUMBER: OPHD-02304
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (415) 705-8410
;              TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 68:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 451 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-271-012-68

  Query Match             100.0%;  Score 451;  DB 15;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 4
US-10-729-122-68
; Sequence 68, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;         APPLICANT: Williams, James A.
;                    Thalley, Bruce S.
;         TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                             Botulinum Neurotoxin
;         NUMBER OF SEQUENCES: 82
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Medlen & Carroll
;              STREET: 220 Montgomery Street, Suite 2200
;              CITY: San Francisco
;              STATE: California
;              COUNTRY: United States of America
;              ZIP: 94104
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/729,122
;              FILING DATE: 05-Dec-2003
;              CLASSIFICATION: 424
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US/08/704,159
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Ingolia, Diane E.
;              REGISTRATION NUMBER: 40,027
;              REFERENCE/DOCKET NUMBER: OPHD-02304
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (415) 705-8410
;              TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 68:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 451 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-729-122-68

  Query Match             100.0%;  Score 451;  DB 16;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 5
US-10-729-039-68
```

```
; Sequence 68, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 68:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-729-039-68

  Query Match             100.0%;  Score 451;  DB 16;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 6
US-10-272-898-66
; Sequence 66, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 66:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1276 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-272-898-66

  Query Match             94.2%;  Score 425;  DB 12;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWVE 451
                 |||||
Db       1272 PGWVE 1276


RESULT 7
US-10-354-774-66
; Sequence 66, Application US/10354774
; Publication No. US20030215468A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 66:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1276 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-354-774-66

  Query Match             94.2%;  Score 425;  DB 15;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971
```

```
Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWVE 451
                 |||||
Db       1272 PGWVE 1276


RESULT 8
US-10-271-012-66
; Sequence 66, Application US/10271012
; Publication No. US20030219457A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/271,012
;            FILING DATE: 15-Oct-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 66:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1276 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-271-012-66

  Query Match             94.2%;  Score 425;  DB 15;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy          87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy         147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy         207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy         267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy         327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy         387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy         447 PGWVE 451
               |||||
Db        1272 PGWVE 1276
```

RESULT 9
US-10-452-024-164
; Sequence 164, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-164

Query Match 94.2%; Score 425; DB 15; Length 1276;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy          87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy         147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy         207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy         267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy         327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy         387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy         447 PGWVE 451
               |||||
Db        1272 PGWVE 1276
```

RESULT 10
US-10-729-122-66
; Sequence 66, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-729-122-66

Query Match 94.2%; Score 425; DB 16; Length 1276;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911
```

```
Qy          87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy         147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy         207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy         267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy         327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy         387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy         447 PGWVE 451
               |||||
Db        1272 PGWVE 1276
```

RESULT 11
US-10-729-039-66
; Sequence 66, Application US/10729039
; Publication No. US20040142455A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,039
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 1996/08/28
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-729-039-66

```
  Query Match           94.2%;  Score 425;  DB 16;  Length 1276;
  Best Local Similarity 100.0%;  Pred. No. 0;
  Matches  425;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy          87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy         147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy         207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy         267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy         327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy         387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy         447 PGWVE 451
               |||||
Db        1272 PGWVE 1276
```

RESULT 12
US-10-452-024-4
; Sequence 4, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-4

```
  Query Match           71.8%;  Score 324;  DB 15;  Length 1276;
  Best Local Similarity 99.8%;  Pred. No. 9.4e-299;
  Matches  424;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy          87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy         147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGN 266
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWVE 451
                |||||
Db       1272 PGWVE 1276


RESULT 13
US-10-452-024-168
; Sequence 168, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 168
;   LENGTH: 1276
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-168

  Query Match             71.8%;  Score 324;  DB 15;  Length 1276;
  Best Local Similarity   99.8%;  Pred. No. 9.4e-299;
  Matches  424;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGN 266
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWVE 451
                |||||
Db       1272 PGWVE 1276


RESULT 14
US-10-205-516-8
; Sequence 8, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 1276
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-8

  Query Match             71.8%;  Score 324;  DB 15;  Length 1276;
  Best Local Similarity   99.8%;  Pred. No. 9.4e-299;
  Matches  424;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGN 266
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWVE 451
                |||||
Db       1272 PGWVE 1276


RESULT 15
US-10-205-516-22
; Sequence 22, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
```

; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 1286
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-22

Query Match 71.8%; Score 324; DB 15; Length 1286;
Best Local Similarity 99.8%; Pred. No. 9.5e-299;
Matches 424; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 915

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 975

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1035

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1095

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1096 PLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILN 1155

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1215

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1275

Qy        447 PGWVE 451
              |||||
Db       1276 PGWVE 1280
```

Search completed: September 10, 2004, 05:52:58
Job time : 27.8137 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 3.89781 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 451
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	324	71.8	1276	1	BXD_CLOBO	P19321 clostridium
2	11	2.4	303	1	MOX2_HUMAN	P50222 homo sapien
3	11	2.4	303	1	MOX2_MOUSE	P32443 mus musculu
4	11	2.4	303	1	MOX2_RAT	P39020 rattus norv
5	11	2.4	354	1	OTX1_HUMAN	P32242 homo sapien
6	11	2.4	355	1	OTX1_MOUSE	P80205 mus musculu
7	11	2.4	355	1	OTX1_RAT	Q63410 rattus norv
8	11	2.4	414	1	TYY1_HUMAN	P25490 homo sapien
9	11	2.4	414	1	TYY1_MOUSE	Q00899 mus musculu
10	11	2.4	420	1	YBE1_SCHPO	O42980 schizosacch
11	11	2.4	466	1	ZIC3_MOUSE	Q62521 mus musculu
12	11	2.4	467	1	ZIC3_HUMAN	O60481 homo sapien
13	11	2.4	611	1	SNF1_CANGA	Q00372 candida gla
14	11	2.4	619	1	SNF1_CANTR	O94168 candida tro
15	11	2.4	1290	1	BXC1_CLOBO	P18640 clostridium
16	10	2.2	202	1	HAN1_CHICK	Q90691 gallus gall
17	10	2.2	204	1	HAN1_SHEEP	Q28555 ovis aries
18	10	2.2	298	1	MOX2_XENLA	P39021 xenopus lae
19	10	2.2	331	1	HXA1_MOUSE	P09022 mus musculu
20	10	2.2	335	1	HXA1_HUMAN	P49639 homo sapien
21	10	2.2	351	1	HRPX_PLALO	P04929 plasmodium
22	10	2.2	375	1	CAR2_DICDI	P34907 dictyosteli
23	10	2.2	387	1	QIN_AVIS3	P56260 avian sarco
24	10	2.2	390	1	PC_DROME	P26017 drosophila
25	10	2.2	410	1	PO42_HUMAN	Q12837 homo sapien
26	10	2.2	411	1	PO42_MOUSE	Q63934 mus musculu
27	10	2.2	430	1	NOR2_RAT	Q63516 rattus norv
28	10	2.2	449	1	GAT6_HUMAN	Q92908 homo sapien
29	10	2.2	451	1	FXGB_CHICK	Q90964 gallus gall
30	10	2.2	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
31	10	2.2	465	1	HNF6_MOUSE	O08755 mus musculu
32	10	2.2	465	1	HNF6_RAT	P70512 rattus norv
33	10	2.2	469	1	FXGA_HUMAN	P55316 homo sapien
34	10	2.2	477	1	FXGB_HUMAN	P55315 homo sapien
35	10	2.2	480	1	FXGB_RAT	Q00939 rattus norv
36	10	2.2	481	1	FXGB_MOUSE	Q60987 mus musculu
37	10	2.2	551	1	CBX4_MOUSE	O55187 mus musculu
38	10	2.2	558	1	CBX4_HUMAN	O00257 homo sapien
39	10	2.2	620	1	SNF1_CANAL	P52497 candida alb
40	10	2.2	623	1	PNT1_DROME	P51022 drosophila
41	10	2.2	626	1	NR43_HUMAN	Q92570 homo sapien
42	10	2.2	628	1	NR43_RAT	P51179 rattus norv
43	10	2.2	633	1	SNF1_YEAST	P06782 saccharomyc
44	10	2.2	657	1	KNOB_PLAFN	P06719 plasmodium
45	10	2.2	665	1	LAMA_XENLA	P11048 xenopus lae

ALIGNMENTS

RESULT 1
BXD_CLOBO
ID BXD_CLOBO STANDARD; PRT; 1276 AA.
AC P19321;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE (Bontoxilysin D).
GN BOTD.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BVD/-3;
RX MEDLINE=91016853; PubMed=2216736;
RA Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT "Nucleotide sequence of the gene encoding Clostridium botulinum
RT neurotoxin type D.";
RL Nucleic Acids Res. 18:5556-5556(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CB16;
RX MEDLINE=93042276; PubMed=1420572;
RA Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT "The complete amino acid sequence of the Clostridium botulinum type D
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT phage d-16 phi genome.";
RL J. Vet. Med. Sci. 54:905-913(1992).
RN [3]
RP PARTIAL SEQUENCE.
RC STRAIN=D-SA, and D-1873;
RX MEDLINE=89339741; PubMed=2668193;
RA Moriishi K., Syuto B., Kubo S., Oguma K.;
RT "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT strains.";
RL Infect. Immun. 57:2886-2891(1989).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinal neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC SYNAPTOBREVINS-1 AND -2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

```
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             71.8%;  Score 324;  DB 1;  Length 1276;
  Best Local Similarity   99.8%;  Pred. No. 2.6e-302;
  Matches  424;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWVE 451
              |||||
Db       1272 PGWVE 1276


RESULT 2
MOX2_HUMAN
ID   MOX2_HUMAN     STANDARD;      PRT;   303 AA.
AC   P50222; Q9UPL6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE   homeobox).
GN   MEOX2 OR MOX2 OR GAX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95331791; PubMed=7607679;
RA   Grigoriou M., Kastrinaki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA   Pachnis V., Karagogeos D.;
RT   "Isolation of the human MOX2 homeobox gene and localization to
RT   chromosome 7p22.1-p21.3.";
RL   Genomics 26:550-555(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95229154; PubMed=7713505;
RA   Lepage D.F., Walsh K.;
RT   "Molecular cloning and localization of the human GAX gene to 7p21.";
RL   Genomics 24:535-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
```

```
RP   SEQUENCE OF 230-303 FROM N.A.
RA   Cordes M., Lacy M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Embryo and placenta.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82629; CAA57949.1; -.
DR   EMBL; L36328; AAA58497.1; -.
DR   EMBL; BC017021; AAH17021.1; -.
DR   EMBL; AC004452; AAC06184.1; -.
DR   PIR; A55641; A55641.
DR   PIR; A56837; A56837.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04005; -.
DR   Genew; HGNC:7014; MEOX2.
DR   MIM; 600535; -.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DNA_BIND    186    245       HOMEOBOX.
FT   CONFLICT     58     58       G -> D (IN REF. 2).
FT   CONFLICT     79     79       MISSING (IN REF. 2).
SQ   SEQUENCE   303 AA;  33457 MW;  809ADE0CD090023D CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0087;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 3
MOX2_MOUSE
ID   MOX2_MOUSE     STANDARD;      PRT;   303 AA.
AC   P32443;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201999; PubMed=1363541;
RA   Candia A.F., Hu J., Crosby J., Lalley P.A., Noden D., Nadeau J.H.,
RA   Wright C.V.E.;
RT   "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are
RT   differentially expressed during early mesodermal patterning in mouse
RT   embryos.";
RL   Development 116:1123-1136(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94232829; PubMed=7909944;
RA   Candia A.F., Kovalik J.-P., Wright C.V.E.;
RT   "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat
RT   homologs.";
RL   Nucleic Acids Res. 21:4982-4982(1993).
RN   [3]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=95349593; PubMed=7623821;
RA   Andres V., Fisher S., Wearsch P., Walsh K.;
RT   "Regulation of Gax homeobox gene transcription by a combination of
RT   positive factors including myocyte-specific enhancer factor 2.";
RL   Mol. Cell. Biol. 15:4272-4281(1995).
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-
CC       8.5 dpc it is found on the entire epithelium of the somite. At 9.5
CC       dpc its expression is restricted to the sclerotome. At 10.5 dpc it
CC       is found in sclerotomally derived cells including the vertebral
CC       and costal precursors.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16406; CAA78899.1; -.
DR   EMBL; S79168; -; NOT_ANNOTATED_CDS.
DR   PIR; B49122; B49122.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04048; -.
DR   MGD; MGI:103219; Meox2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       63     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33506 MW;  41BD05FC39AA4427 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0087;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77
```

```
RESULT 4
MOX2_RAT
ID   MOX2_RAT       STANDARD;      PRT;   303 AA.
AC   P39020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Growth arrest-specific homeobox).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93268321; PubMed=8098844;
RA   Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,
RA   Walsh K.;
RT   "Molecular cloning of a diverged homeobox gene that is rapidly down-
RT   regulated during the G0/G1 transition in vascular smooth muscle
RT   cells.";
RL   Mol. Cell. Biol. 13:3722-3733(1993).
RN   [2]
RP   REVISIONS.
RA   Walsh K.;
RL   Submitted (MAR-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Aorta and heart. Also detected in lung and
CC       kidney.
CC   -!- INDUCTION: Rapidly and transiently down-regulated during the
CC       transition from G0 to G1 induced by mitogen stimulation.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z17223; CAA78931.1; -.
DR   PIR; A48130; A48130.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04054; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       64     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33605 MW;  7776642AEFA3A2E8 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0087;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
OTX1_HUMAN
ID   OTX1_HUMAN     STANDARD;      PRT;   354 AA.
AC   P32242;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS):
CC       5'-TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC007621; AAH07621.1; -.
DR   PIR; S39406; S39406.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02079; -.
DR   Genew; HGNC:8521; OTX1.
```

```
DR   MIM; 600036; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    301       HIS-RICH.
SQ   SEQUENCE   354 AA;  37327 MW;  E32C1E47464BDDAA CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 6
OTX1_MOUSE
ID   OTX1_MOUSE     STANDARD;      PRT;   355 AA.
AC   P80205;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1 OR OTX-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 31-98 FROM N.A.
RX   MEDLINE=92357147; PubMed=1353865;
RA   Simeone A., Acampora D., Gulisano M., Stornaiuolo A., Boncinelli E.;
RT   "Nested expression domains of four homeobox genes in developing
RT   rostral brain.";
RL   Nature 358:687-690(1992).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC057105; AAH57105.1; -.
DR   EMBL; BC058354; AAH58354.1; -.
DR   EMBL; X68883; CAA48754.1; -.
DR   PIR; S35345; S35345.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02080; -.
DR   MGD; MGI:97450; Otx1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
SQ   SEQUENCE   355 AA;  37531 MW;  DA5136D9D4FC948F CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 7
OTX1_RAT
ID   OTX1_RAT       STANDARD;      PRT;   355 AA.
AC   Q63410; Q64203;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95016961; PubMed=7931541;
RA   Frantz G.D., Weimann J.M., Levin M.E., McConnell S.K.;
RT   "Otx1 and Otx2 define layers and regions in developing cerebral
RT   cortex and cerebellum.";
RL   J. Neurosci. 14:5725-5740(1994).
RN   [2]
RP   SEQUENCE OF 176-323 FROM N.A.
RX   MEDLINE=96108898; PubMed=8613727;
RA   Robel L., Ding M., James A.J., Lin X., Simeone A., Leckman J.F.,
RA   Vaccarino F.M.;
RT   "Fibroblast growth factor 2 increases Otx2 expression in precursor
RT   cells from mammalian telencephalon.";
RL   J. Neurosci. 15:7879-7891(1995).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3' (By similarity). May play a role in the specification
CC       or differentiation of neurons in the deep layers of the cerebral
CC       cortex, and also in cerebellar regionalization during early
CC       development.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: Expressed throughout the forebrain and
CC       midbrain during development, and in addition is also seen in
CC       discrete spatial and temporal domains in the developing cerebral
CC       cortex and cerebellum. Confined to a subpopulation of neurons in
CC       layers 5 and 6 within the adult cerebral cortex and during
CC       development expression is high in the progenitors of these deep-
CC       layer cells. Expressed in the developing cerebellum in spatially
CC       restricted regions of the external granular layer.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32602; AAA53557.1; -.
DR   EMBL; S81924; -; NOT_ANNOTATED_CDS.
DR   PIR; I56547; I56547.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND    38     97       HOMEOBOX.
FT   DOMAIN      275    302      HIS-RICH.
FT   CONFLICT    219    219      T -> A (IN REF. 2).
FT   CONFLICT    229    229      R -> G (IN REF. 2).
SQ   SEQUENCE   355 AA;  37602 MW;  C875871723D0B876 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
TYY1_HUMAN
ID   TYY1_HUMAN     STANDARD;      PRT;   414 AA.
AC   P25490; Q14935;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1).
GN   YY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92005716; PubMed=1655281;
RA   Shi Y., Seto E., Chang L.-S., Shenk T.;
RT   "Transcriptional repression by YY1, a human GLI-Kruppel-related
RT   protein, and relief of repression by adenovirus E1A protein.";
RL   Cell 67:377-388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=92052179; PubMed=1946405;
RA   Park K., Atchison M.;
RT   "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
RT   that binds to the immunoglobulin kappa 3' enhancer and the
RT   immunoglobulin heavy-chain mu E1 site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Whitson R.H., Huang T., Dang J., Itakura K.;
RL   Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98152931; PubMed=9493912;
RA   McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
RA   Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
RT   "Targeting of the YY1 transcription factor to the nucleolus and the
RT   nuclear matrix in situ: the C-terminus is a principal determinant for
RT   nuclear trafficking.";
```

```
RL   J. Cell. Biochem. 68:500-510(1998).
RN   [6]
RP   INTERACTION WITH YAF2.
RX   MEDLINE=97169297; PubMed=9016636;
RA   Kalenik J.L., Chen D., Bradley M.E., Chen S.-J., Lee T.-C.;
RT   "Yeast two-hybrid cloning of a novel zinc finger protein that
RT   interacts with the multifunctional transcription factor YY1.";
RL   Nucleic Acids Res. 25:843-849(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
RX   MEDLINE=97098436; PubMed=8942976;
RA   Houbaviy H.B., Usheva A., Shenk T., Burley S.K.;
RT   "Cocrystal structure of YY1 bound to the adeno-associated virus P5
RT   initiator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 353-379.
RX   MEDLINE=98308000; PubMed=9642075;
RA   Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
RA   Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
RT   "Design, synthesis and structure of a zinc finger with an artificial
RT   beta-turn.";
RL   J. Mol. Biol. 279:973-986(1998).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
CC       EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
CC       PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77698; AAA59467.1; -.
DR   EMBL; M76541; AAA59926.1; -.
DR   EMBL; Z14077; CAA78455.1; -.
DR   EMBL; BC037308; AAH37308.1; -.
DR   PIR; A40350; A40350.
DR   PDB; 1UBD; 23-DEC-96.
DR   PDB; 1ZNM; 01-APR-98.
DR   TRANSFAC; T00915; -.
DR   Genew; HGNC:12856; YY1.
DR   MIM; 600013; -.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     69       GLY-RICH.
FT   DOMAIN       70     80       POLY-HIS.
FT   DOMAIN      159    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE 1.
FT   ZN_FING     325    347       C2H2-TYPE 2.
FT   ZN_FING     353    377       C2H2-TYPE 3.
FT   ZN_FING     383    407       C2H2-TYPE 4.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION.
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN.
FT   CONFLICT     65     65       H -> R (IN REF. 2).
FT   CONFLICT    196    196       G -> R (IN REF. 1).
FT   STRAND      296    297
FT   TURN        301    302
FT   STRAND      306    307
FT   HELIX       310    317
FT   HELIX       318    320
FT   STRAND      325    326
FT   TURN        328    330
FT   STRAND      333    334
FT   HELIX       337    343
FT   HELIX       344    346
FT   TURN        347    347
FT   STRAND      353    354
FT   TURN        358    359
FT   STRAND      363    364
FT   HELIX       367    378
FT   TURN        388    389
FT   TURN        397    400
FT   HELIX       401    407
SQ   SEQUENCE   414 AA;  44712 MW;  058C05A0AD2D04E6 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 9
TYY1_MOUSE
ID   TYY1_MOUSE     STANDARD;      PRT;   414 AA.
AC   Q00899;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1) (UCR-motif DNA-binding protein).
GN   YY1 OR UCRBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107191; PubMed=1309593;
RA   Flanagan J.R., Becker K.G., Ennist D.L., Gleason S.L., Driggers P.H.,
RA   Levi B.-Z., Appella E., Ozato K.;
RT   "Cloning of a negative transcription factor that binds to the
RT   upstream conserved region of Moloney murine leukemia virus.";
RL   Mol. Cell. Biol. 12:38-44(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93296177; PubMed=8516301;
RA   Safrany G., Perry R.P.;
RT   "Characterization of the mouse gene that encodes the delta/YY1/NF-
RT   E1/UCRBP transcription factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5559-5563(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=92052178; PubMed=1946404;
RA   Hariharan N., Kelley D.E., Perry R.P.;
RT   "Delta, a transcription factor that binds to downstream elements in
RT   several polymerase II promoters, is a functionally versatile zinc
RT   finger protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9799-9803(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. BINDS TO
CC       THE UPSTREAM CONSERVED REGION (UCR) (5'-CGCCATTTT-3') OF MOLONEY
CC       MURINE LEUKEMIA VIRUS (MULV).
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73963; AAA40522.1; -.
DR   EMBL; L13968; AAA40477.1; -.
DR   EMBL; L13969; AAA40477.1; JOINED.
DR   EMBL; L13965; AAA40477.1; JOINED.
DR   EMBL; L13966; AAA40477.1; JOINED.
DR   EMBL; L13967; AAA40477.1; JOINED.
DR   EMBL; M74590; AAA37521.1; -.
DR   EMBL; BC055899; AAH55899.1; -.
DR   PIR; A48273; A48273.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T00278; -.
DR   TRANSFAC; T00865; -.
DR   MGD; MGI:99150; Yy1.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003700; F:transcription factor activity; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     70       GLY-RICH.
FT   DOMAIN       71     82       POLY-HIS.
FT   DOMAIN      161    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    407       C2H2-TYPE.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION
FT                                (BY SIMILARITY).
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION (BY SIMILARITY).
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN (BY SIMILARITY).
FT   CONFLICT    219    219       F -> S (IN REF. 3).
FT   CONFLICT    375    375       R -> G (IN REF. 3).
SQ   SEQUENCE   414 AA;  44717 MW;  C012378288E984F9 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 10
YBE1_SCHPO
ID   YBE1_SCHPO     STANDARD;      PRT;   420 AA.
AC   O42980; P78807;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative thiol protease C17D11.01 (EC 3.4.22.-).
GN   SPBC17D11.01 OR SPBC20F10.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
```

```
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- SIMILARITY: Belongs to peptidase family C48.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031739; CAA21072.1; -.
DR   EMBL; AL021747; CAA16851.1; -.
DR   EMBL; D89156; BAA13818.1; -.
DR   GeneDB_SPombe; SPBC17D11.01; -.
DR   InterPro; IPR003653; Peptidase_C48.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Thiol protease.
FT   DOMAIN      299    312       POLY-HIS.
FT   CONFLICT     57     57       L -> W (IN REF. 2).
FT   CONFLICT    420    420       F -> FGCYTCVV (IN REF. 2).
SQ   SEQUENCE   420 AA;  47257 MW;  2125D94D9A2D9E87 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 11
ZIC3_MOUSE
ID   ZIC3_MOUSE     STANDARD;      PRT;   466 AA.
AC   Q62521;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=96132843; PubMed=8557628;
RA   Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA   Chapman V.M., Mikoshiba K.;
RT   "The mouse zic gene family. Homologues of the Drosophila pair-rule
RT   gene odd-paired.";
RL   J. Biol. Chem. 271:1043-1047(1996).
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: CNS. A high level expression is seen in the
CC       cerebellum.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D70849; BAA11116.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T04671; -.
DR   MGD; MGI:106676; Zic3.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     96       POLY-HIS.
FT   ZN_FING     294    321       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     327    351       C2H2-TYPE 2.
FT   ZN_FING     357    381       C2H2-TYPE 3.
FT   ZN_FING     387    409       C2H2-TYPE 4.
SQ   SEQUENCE   466 AA;  50644 MW;  089749DF726F069B CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 12
ZIC3_HUMAN
ID   ZIC3_HUMAN     STANDARD;      PRT;   467 AA.
AC   O60481;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HTX1 MET-323.
RX   MEDLINE=98016414; PubMed=9354794;
RA   Gebbia M., Ferrero G.B., Pilia G., Bassi M.T., Aylsworth A.S.,
RA   Penman-Splitt M., Bird L.M., Bamforth J.S., Burn J., Schlessiner D.,
RA   Nelson D.L., Casey B.;
RT   "X-linked situs abnormalities result from mutations in ZIC3.";
RL   Nat. Genet. 17:305-308(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in ZIC3 are the cause of X-Linked visceral
CC       heterotaxy (HTX1) [MIM:306955]. HTX1 is a disease characterized by
CC       congenital heart disease and alterations of visceral situs,
CC       including asplenia or polysplenia, symmetric liver, intestinal
CC       malrotation, and abnormal lung lobation.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
```

```
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028706; AAC05594.1; -.
DR   EMBL; AL035443; CAB41648.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:12874; ZIC3.
DR   MIM; 300265; -.
DR   MIM; 306955; -.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein; Disease mutation.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     97.      POLY-HIS.
FT   ZN_FING     295    322       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     328    352       C2H2-TYPE 2.
FT   ZN_FING     358    382       C2H2-TYPE 3.
FT   ZN_FING     388    410       C2H2-TYPE 4.
FT   VARIANT     323    323       T -> M (in HTX1).
FT                                /FTId=VAR_007753.
SQ   SEQUENCE   467 AA;  50569 MW;  3150CF13C0679568 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 13
SNF1_CANGA
ID   SNF1_CANGA     STANDARD;      PRT;   611 AA.
AC   Q00372;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCCLS84;
RX   MEDLINE=97101049; PubMed=8945576;
RA   Petter R., Kwon-Chung K.J.;
RT   "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT   pathogenic yeast Candida glabrata.";
RL   Infect. Immun. 64:5269-5273(1996).
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78130; AAB48642.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN        6     17       POLY-HIS.
FT   DOMAIN       39    290       PROTEIN KINASE.
FT   NP_BIND      45     53       ATP (BY SIMILARITY).
FT   BINDING      68     68       ATP (BY SIMILARITY).
FT   ACT_SITE    161    161       BY SIMILARITY.
FT   MOD_RES     194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  70049 MW;  89E17812A4900CD0 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 0.016;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db          8 HHHHHHHHHHS 18


RESULT 14
SNF1_CANTR
ID   SNF1_CANTR     STANDARD;      PRT;   619 AA.
AC   O94168;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanai T., Ogawa K., Ueda M., Tanaka A.;
RT   "Genetic evaluation of the function of SNF1 in Candida tropicalis.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB024535; BAA75889.1; -.
```

```
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN       17     29       POLY-HIS.
FT   DOMAIN       52    303       PROTEIN KINASE.
FT   NP_BIND      58     66       ATP (BY SIMILARITY).
FT   BINDING      81     81       ATP (BY SIMILARITY).
FT   ACT_SITE    174    174       BY SIMILARITY.
FT   MOD_RES     207    207       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  70323 MW;  0FCF1FC3DCE706D7 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 0.016;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db         20 HHHHHHHHHHS 30


RESULT 15
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             2.4%;  Score 11;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFN 37
              ||||||||||
Db        855 LLKDIINEYFN 865


Search completed: September 10, 2004, 05:07:21
Job time : 3.89781 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 18.3427 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 451
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	323	71.6	1275	12	Q9QTG7	Q9qtg7 clostridium
2	80	17.7	1280	2	Q45849	Q45849 clostridium
3	80	17.7	1280	2	Q9LBS7	Q9lbs7 clostridium
4	80	17.7	1280	2	Q841S3	Q841s3 clostridium
5	21	4.7	1285	2	Q45967	Q45967 clostridium
6	21	4.7	1285	2	Q9LBR1	Q9lbr1 clostridium
7	12	2.7	226	11	Q9QZD2	Q9qzd2 rattus norv
8	12	2.7	516	13	O42186	O42186 brachydanio
9	12	2.7	517	13	Q90YL6	Q90yl6 brachydanio
10	12	2.7	569	13	Q802D2	Q802d2 brachydanio
11	12	2.7	679	13	Q803K9	Q803k9 brachydanio
12	12	2.7	706	5	Q8SXG1	Q8sxg1 drosophila
13	12	2.7	861	5	Q9VRT8	Q9vrt8 drosophila
14	11	2.4	163	5	Q7YTA3	Q7yta3 glomeris ma
15	11	2.4	225	13	Q90XZ3	Q90xz3 erpetoichth
16	11	2.4	230	4	O75263	O75263 homo sapien
17	11	2.4	243	4	Q9P2R2	Q9p2r2 homo sapien
18	11	2.4	243	6	Q9N2G2	Q9n2g2 gorilla gor
19	11	2.4	243	6	Q9N2G3	Q9n2g3 pan troglod
20	11	2.4	243	6	Q9N2G1	Q9n2g1 pongo pygma
21	11	2.4	302	13	Q90YH7	Q90yh7 gallus gall
22	11	2.4	303	6	Q95JA6	Q95ja6 sus scrofa
23	11	2.4	303	11	Q99M23	Q99m23 mus musculu
24	11	2.4	355	11	Q91ZJ3	Q91zj3 mus musculu
25	11	2.4	358	13	Q9DG32	Q9dg32 xenopus lae
26	11	2.4	411	11	Q8CHJ4	Q8chj4 rattus norv
27	11	2.4	441	5	Q8T1W3	Q8t1w3 dictyosteli
28	11	2.4	449	10	Q8L4M9	Q8l4m9 oryza sativ
29	11	2.4	456	11	Q8BSB3	Q8bsb3 mus musculu
30	11	2.4	467	13	Q9YGC6	Q9ygc6 xenopus lae
31	11	2.4	511	10	Q8GXZ5	Q8gxz5 arabidopsis
32	11	2.4	906	5	Q8IN91	Q8in91 drosophila
33	11	2.4	1122	5	Q8SXD4	Q8sxd4 drosophila
34	11	2.4	1147	5	Q95PH8	Q95ph8 dictyosteli
35	11	2.4	1180	5	Q24163	Q24163 drosophila
36	11	2.4	1186	5	Q9VED3	Q9ved3 drosophila
37	11	2.4	1283	5	Q95RH4	Q95rh4 drosophila
38	11	2.4	1291	2	Q93HT3	Q93ht3 clostridium
39	11	2.4	1911	5	Q9W4M7	Q9w4m7 drosophila
40	10	2.2	53	2	O55244	O55244 escherichia
41	10	2.2	233	4	Q8TC53	Q8tc53 homo sapien
42	10	2.2	233	5	Q8MP30	Q8mp30 dictyosteli
43	10	2.2	236	11	Q9QWE1	Q9qwe1 mus sp. cla
44	10	2.2	239	5	Q7Z1G6	Q7z1g6 trypanosoma
45	10	2.2	244	13	Q8JHA3	Q8jha3 gallus gall

ALIGNMENTS

RESULT 1
Q9QTG7
ID Q9QTG7 PRELIMINARY; PRT; 1275 AA.
AC Q9QTG7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NTX (Fragment).
GN NTX.
OS Clostridium botulinum D bacteriophage.
OC Viruses.
OX NCBI_TaxID=29342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1873;
RX MEDLINE=99017546; PubMed=9802560;
RA Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT "Molecular composition of the 16S toxin produced by a Clostridium
RT botulinum type D strain, 1873.";
RL Microbiol. Immunol. 42:599-605(1998).
DR EMBL; AB012112; BAA75084.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1275 1275
SQ SEQUENCE 1275 AA; 146742 MW; 3C50F46C8233E2D6 CRC64;

```
  Query Match            71.6%;  Score 323;  DB 12;  Length 1275;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches  423;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 911

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNI 971

Qy        147 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 EWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIED 1031

Qy        207 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGN 1091

Qy        267 PLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILN 326
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1092 PLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILN 1151

Qy        327 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 GDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIV 1211

Qy        387 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 SKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRD 1271

Qy        447 PGWV 450
              ||||
Db       1272 PGWV 1275
```

```
RESULT 2
Q45849
ID   Q45849      PRELIMINARY;      PRT;  1280 AA.
AC   Q45849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum C.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=36828;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6813;
RX   MEDLINE=96156810; PubMed=8593068;
RA   Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA   Oguma K.;
RT   "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT   composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT   detection of this gene from Clostridium botulinum type C organisms.";
RL   Appl. Environ. Microbiol. 62:662-667(1996).
DR   EMBL; D49440; BAA08418.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147815 MW;  74F813B228B8C989 CRC64;

  Query Match            17.7%;  Score 80;  DB 2;  Length 1280;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-74;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        209 EVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPL 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 EVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPL 1097

Qy        269 KFDTEYYIINDNYIDRYIAP 288
              ||||||||||||||||||||
Db       1098 KFDTEYYIINDNYIDRYIAP 1117
```

```
RESULT 3
Q9LBS7
ID   Q9LBS7      PRELIMINARY;      PRT;  1280 AA.
AC   Q9LBS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-6814;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Kawabe T.,
RA   Murakami F., Nakatsuka M., Ohyama T.;
RT   "Organization of Gene Encoding Components of the Botulinum Progenitor
RT   Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT   Chimeric Sequence in the Gene Encoding Each Component.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037166; BAA89713.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147757 MW;  DE124FFB6F68450B CRC64;

  Query Match            17.7%;  Score 80;  DB 2;  Length 1280;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-74;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        209 EVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPL 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 EVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPL 1097

Qy        269 KFDTEYYIINDNYIDRYIAP 288
              ||||||||||||||||||||
Db       1098 KFDTEYYIINDNYIDRYIAP 1117
```

```
RESULT 4
Q841S3
ID   Q841S3      PRELIMINARY;      PRT;  1280 AA.
AC   Q841S3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2003;
RA   Lee M.S., Cheng M.C., Chen Y.P., Shiau J.R., Lin S.Y.;
RT   "An outbreak of botulism in black-faced spoonbill in Taiwan.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY251553; AAP06952.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1280 AA;  147642 MW;  34C1A2E573EE0AE7 CRC64;

  Query Match             17.7%;  Score 80;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-74;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        209 EVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPL 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 EVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPL 1097

Qy        269 KFDTEYYIINDNYIDRYIAP 288
              ||||||||||||||||||||
Db       1098 KFDTEYYIINDNYIDRYIAP 1117


RESULT 5
Q45967
ID   Q45967      PRELIMINARY;      PRT;  1285 AA.
AC   Q45967;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE D;
RX   MEDLINE=96283801; PubMed=8679691;
RA   Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA   Ogumad K.;
RT   "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT   types C and D organisms.";
RL   Biochim. Biophys. Acta 1307:123-126(1996).
DR   EMBL; D38442; BAA07477.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1285 AA;  147365 MW;  EEDE98E4ECA6C413 CRC64;

  Query Match             4.7%;  Score 21;  DB 2;  Length 1285;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         32 INEYFNSINDSKILSLQNKKN 52
              ||||||||||||||||||||
Db        857 INEYFNSINDSKILSLQNKKN 877


RESULT 6
Q9LBR1
ID   Q9LBR1      PRELIMINARY;      PRT;  1285 AA.
AC   Q9LBR1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-4947;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA   Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT   "Characterization of the Progenitor Toxin Components Produced by
RT   Clostridium botulinum Type D Strain 4947.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037920; BAA90661.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1285 AA;  147352 MW;  B63AFA487D570680 CRC64;

  Query Match             4.7%;  Score 21;  DB 2;  Length 1285;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         32 INEYFNSINDSKILSLQNKKN 52
```

```
                    ||||||||||||||||||||
Db        857 INEYFNSINDSKILSLQNKKN 877


RESULT 7
Q9QZD2
ID   Q9QZD2      PRELIMINARY;      PRT;   226 AA.
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
SQ   SEQUENCE   226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

  Query Match             2.7%;  Score 12;  DB 11;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 0.00097;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db         17 HHHHHHHHHHSS 28


RESULT 8
O42186
ID   O42186      PRELIMINARY;      PRT;   516 AA.
AC   O42186;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FTZ-F1 homolog.
GN   NR5A2 OR ZFF1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322094; PubMed=9178748;
RA   Liu D., Le Drean Y., Ekker M., Xiong F., Hew C.L.;
RT   "Teleost FTZ-F1 homolog and its splicing variant determine the
RT   expression of the salmon gonadotropin IIbeta subunit gene.";
RL   Mol. Endocrinol. 11:877-890(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AF014926; AAC60274.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04065; -.
DR   ZFIN; ZDB-GENE-990415-79; nr5a2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   516 AA;  58805 MW;  500DA09E116D5FA5 CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 9
Q90YL6
ID   Q90YL6      PRELIMINARY;      PRT;   517 AA.
AC   Q90YL6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nuclear receptor FF1aE.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chai C., Tan J.H., Gao W., Teh H.L., Chan W.K.;
RT   "Full length zebrafish ff1aE cDNA.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AY034061; AAK54449.1; -.
DR   HSSP; P19793; 1LBD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   517 AA;  58761 MW;  CF48D08AB7A75F29 CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200
```

```
RESULT 10
Q802D2
ID   Q802D2      PRELIMINARY;      PRT;   569 AA.
AC   Q802D2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048886; AAH48886.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
KW   Cyclin.
SQ   SEQUENCE   569 AA;  64117 MW;  812A7332C06EF4C2 CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 0.0023;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        486 HHHHHHHHHHSS 497


RESULT 11
Q803K9
ID   Q803K9      PRELIMINARY;      PRT;   679 AA.
AC   Q803K9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T1 (Fragment).
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044435; AAH44435.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
KW   Cyclin.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  74887 MW;  AB5667D59439659F CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 679;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        488 HHHHHHHHHHSS 499


RESULT 12
Q8SXG1
ID   Q8SXG1      PRELIMINARY;      PRT;   706 AA.
AC   Q8SXG1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RH12258p.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AY089654; AAL90392.1; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   706 AA;  79683 MW;  73C0F8199DF60E5F CRC64;

  Query Match             2.7%;  Score 12;  DB 5;  Length 706;
  Best Local Similarity   100.0%;  Pred. No. 0.0028;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        370 GHHHHHHHHHHS 381


RESULT 13
Q9VRT8
ID   Q9VRT8      PRELIMINARY;      PRT;   861 AA.
AC   Q9VRT8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG6619 protein.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
```

```
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AE003564; AAF50702.2; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   861 AA;  96809 MW;  8530ED5D92671A41 CRC64;

  Query Match             2.7%;  Score 12;  DB 5;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 0.0033;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        525 GHHHHHHHHHHS 536


RESULT 14
Q7YTA3
ID   Q7YTA3       PRELIMINARY;      PRT;   163 AA.
AC   Q7YTA3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Distal-less protein (Fragment).
GN   DLL.
OS   Glomeris marginata.
OC   Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Pentazonia;
OC   Glomerida; Glomeridae; Glomeris.
OX   NCBI_TaxID=62006;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22767835; PubMed=12885558;
RA   Prpic N.M., Tautz D.;
RT   "The expression of the proximodistal axis patterning genes Distal-less
RT   and dach shund in the appendages of Glomeris marginata (Myriapoda:
RT   Diplopoda) suggests a special role for these genes in patterning the
RT   head appendages.";
RL   Dev. Biol. 260:97-112(2003).
DR   EMBL; AJ551276; CAD82905.1; -.
FT   NON_TER       1      1
FT   NON_TER     163    163
SQ   SEQUENCE   163 AA;  17588 MW;  87B8BCD9F5421697 CRC64;

  Query Match             2.4%;  Score 11;  DB 5;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 0.0081;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHH 11
              |||||||||||
Db         16 MGHHHHHHHHH 26


RESULT 15
Q90XZ3
ID   Q90XZ3       PRELIMINARY;      PRT;   225 AA.
AC   Q90XZ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Homeodomain protein Otx1 (Fragment).
OS   Erpetoichthys calabaricus (Rope fish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Polypteriformes; Polypteridae; Erpetoichthys.
OX   NCBI_TaxID=27687;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21396394; PubMed=11504847;
RA   Germot A., Lecointre G., Plouhinec J.L., Le Mentec C., Girardot F.,
RA   Mazan S.;
RT   "Structural evolution of otx genes in Craniates.";
RL   Mol. Biol. Evol. 18:1668-1678(2001).
DR   EMBL; AF321406; AAL01673.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     225    225
SQ   SEQUENCE   225 AA;  23475 MW;  0AE78B8DB48B4941 CRC64;

  Query Match             2.4%;  Score 11;  DB 13;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||||
Db        172 GHHHHHHHHHH 182


Search completed: September 10, 2004, 05:12:56
Job time : 18.3427 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 29.7495 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 2397
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2394	99.9	451	2	AAW68398	Aaw68398 Clostridi
2	2228	92.9	451	4	AAB04093	Aab04093 Botulism
3	963.5	40.2	462	2	AAW68397	Aaw68397 Clostridi
4	801.5	33.4	450	3	AAY77136	Aay77136 Synthetic
5	801.5	33.4	450	4	AAB04092	Aab04092 Botulism
6	801.5	33.4	1291	2	AAY05814	Aay05814 Non-toxic
7	772.5	32.2	472	2	AAW68394	Aaw68394 Clostridi
8	765.5	31.9	472	2	AAW68393	Aaw68393 Clostridi
9	762	31.8	473	2	AAW68400	Aaw68400 Clostridi
10	749	31.2	462	2	AAR95009	Aar95009 Type A ne
11	749	31.2	462	2	AAW68390	Aaw68390 Clostridi
12	698.5	29.1	448	2	AAW68399	Aaw68399 Clostridi
13	693	28.9	452	2	AAW68396	Aaw68396 Clostridi
14	684.5	28.6	451	2	AAW68395	Aaw68395 Clostridi
15	627	26.2	445	2	AAW68391	Aaw68391 Clostridi
16	620.5	25.9	1070	3	AAY93308	Aay93308 A mangane
17	620.5	25.9	1095	3	AAY93311	Aay93311 A mangane
18	619.5	25.8	1291	2	AAW68392	Aaw68392 Clostridi
19	617.5	25.8	439	4	AAB04085	Aab04085 Botulism
20	616.5	25.7	399	3	AAY78982	Aay78982 C. botuli
21	616.5	25.7	440	3	AAY77135	Aay77135 Synthetic
22	616.5	25.7	440	4	AAB04091	Aab04091 Botulism
23	613.5	25.6	848	4	AAB04082	Aab04082 Botulism
24	610	25.4	449	3	AAY77139	Aay77139 Synthetic
25	610	25.4	449	4	AAB04167	Aab04167 Botulism
26	596	24.9	1169	2	AAW56017	Aaw56017 Recombina
27	595	24.8	438	2	AAR95008	Aar95008 Type A ne
28	595	24.8	438	2	AAW68389	Aaw68389 Clostridi
29	595	24.8	438	3	AAY77134	Aay77134 Synthetic
30	588	24.5	437	4	AAB04088	Aab04088 Botulism
31	588	24.5	1067	3	AAY93307	Aay93307 A mangane
32	588	24.5	1092	3	AAY93310	Aay93310 A mangane
33	588	24.5	1296	2	AAR95010	Aar95010 C. botuli
34	587	24.5	434	4	AAB04089	Aab04089 Botulism
35	587	24.5	435	4	AAB04090	Aab04090 Botulism
36	581	24.2	1295	5	AAU99339	Aau99339 Clostridi
37	577	24.1	432	3	AAY77142	Aay77142 Native bo
38	575	24.0	837	3	AAY77140	Aay77140 Native bo
39	564	23.5	847	4	AAB04081	Aab04081 Botulism
40	560	23.4	862	4	AAE07890	Aae07890 Modified
41	560	23.4	887	4	AAE07892	Aae07892 Modified
42	560	23.4	979	6	AAE35713	Aae35713 BoNT/F-Hc
43	560	23.4	1192	6	AAE35711	Aae35711 BoNT/F-Hc
44	560	23.4	1192	6	AAE35710	Aae35710 BoNT/F-Hc
45	559.5	23.3	685	4	AAE07893	Aae07893 Modified

ALIGNMENTS

RESULT 1
AAW68398
ID AAW68398 standard; protein; 451 AA.
XX
AC AAW68398;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type D toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW botulism; BotD.
XX
OS Clostridium botulinum; serotype D.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30591.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 47; Page 352-353; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC BotD soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which

```
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             99.9%;  Score 2394;  DB 2;  Length 451;
  Best Local Similarity   99.8%;  Pred. No. 6.2e-173;
  Matches  450;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNLLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              ||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 2
AAB04093
ID   AAB04093 standard; protein; 451 AA.
XX
AC   AAB04093;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype D).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54487.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 6b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 451 AA;

  Query Match             92.9%;  Score 2228;  DB 4;  Length 451;
  Best Local Similarity   99.3%;  Pred. No. 2.3e-160;
  Matches  423;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 325

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGDCSQNCVYALKLQSNLGNYGIGIFSIKNI 385

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445

Qy        446 DPGWVE 451
              ||||||
Db        446 DPGWVE 451


RESULT 3
AAW68397
```

```
ID   AAW68397 standard; protein; 462 AA.
XX
AC   AAW68397;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type C1 toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW   botulism; BotC.
XX
OS   Clostridium botulinum; serotype C1 strain Stockholm.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30588.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 45; Page 339-341; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC   DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC   express BotC soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             40.2%;  Score 963.5;  DB 2;  Length 462;
  Best Local Similarity   45.0%;  Pred. No. 1.5e-64;
  Matches  209;  Conservative   82;  Mismatches  146;  Indels   27;  Gaps   11;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||:||||||||||:|| |||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNNINDSKILSLQNRKNTLVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGD---KIIVNLNNNILYSAIYENSSVSFWIKISKD 117
              ||||    :|||| |:  |||| |||:   |:||  | ||:|:::||: |:||||:|:|
Db         61 NAEVSEEGDVQLNPIFPPDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKW 120

Qy        118 LTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKW 177
              ::|    ||||:|:: |||| : | :  : : |:      :|: | |  | :  || |||
Db        121 VSNLPG-YTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKW 178

Qy        178 FFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENID--------ENQM 229
              ||||:|||:|| ||:||||:|  : |:::|  :   ||| | |::  |  .      :|

Db        179 FFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNIN 238

Qy        230 LWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPE 289
              :||||| ||:|||  :||||::      ||:|||||| |::: |||::| :|::||:
Db        239 MWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYAN 298

Qy        290 SNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDT 349
              | :|     : :    |  | || :      :|:  || :   |  |:: | :    :|
Db        299 SRQIVFNTRRNNNDFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNET 358

Qy        350 IYATQGGECSQNCVYALKLQSNLGNYGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLL 406
              :||       |   :||: |:    :     || |: : :   | |||| |:|  ||   :
Db        359 MYADNH---STEDIYAIGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGI 415

Qy        407 ADIYKPWRFS-----FKNAYTPVAVT--NYETKLLSTSSFWKFI 443
              |   :||      ::: |     |   || : | |||: | |:
Db        416 CSI-GTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFV 458


RESULT 4
AAY77136
ID   AAY77136 standard; protein; 450 AA.
XX
AC   AAY77136;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype C; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87214.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 24; Page 41-42; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
```

CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 450 AA;

Query Match 33.4%; Score 801.5; DB 3; Length 450;
Best Local Similarity 41.7%; Pred. No. 2.8e-52;
Matches 183; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

```
Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
Db         14 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 73

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
Db         74 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 132

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
Db        133 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 191

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
Db        192 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 251

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
Db        252 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 311

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
Db        312 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 368

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
Db        369 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 427

Qy        427 T--NYETKLLSTSSFWKFI 443
Db        428 KQGNYASLLESTSTHWGFV 446
```

RESULT 5
AAB04092
ID AAB04092 standard; protein; 450 AA.
XX
AC AAB04092;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype C).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54486.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 5b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 450 AA;

Query Match 33.4%; Score 801.5; DB 4; Length 450;
Best Local Similarity 41.7%; Pred. No. 2.8e-52;
Matches 183; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

```
Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
Db         14 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 73

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
Db         74 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 132

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
Db        133 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 191

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
Db        192 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 251

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
Db        252 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 311

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
Db        312 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 368

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
Db        369 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 427

Qy        427 T--NYETKLLSTSSFWKFI 443
Db        428 KQGNYASLLESTSTHWGFV 446
```

RESULT 6

```
AAY05814
ID   AAY05814 standard; protein; 1291 AA.
XX
AC   AAY05814;
XX
DT   02-AUG-1999  (first entry)
XX
DE   Non-toxic modified botulinum toxin rBoNT/C.
XX
KW   Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery; mutant.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 229
FT                   /note= "His in native toxin"
FT   Misc-difference 230
FT                   /note= "Glu in native toxin"
FT   Misc-difference 233
FT                   /note= "His in native toxin"
XX
PN   WO9920306-A1.
XX
PD   29-APR-1999.
XX
PF   16-OCT-1998;   98WO-US021897.
XX
PR   20-OCT-1997;   97US-00954302.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Simpson L,  Kiyatkin N,  Maksymowych A;
XX
DR   WPI; 1999-302646/25.
DR   N-PSDB; AAX25521.
XX
PT   Modified toxin useful for systemic delivery of oral vaccines and
PT   therapeutic agents.
XX
PS   Example 1; Page 31-33; 37pp; English.
XX
CC   The present sequence represents a modified serotype C botulinum toxin,
CC   termed rBoNT/C, in which amino acids His-229, Glu-230 and His-233 of the
CC   native sequence are substituted by Gly, Thr and Asn, respectively, i.e.
CC   the zinc binding motif (see AAY05817) of the light chain holotoxin is
CC   modified, resulting in loss of endoprotease activity. DNA coding for the
CC   modified botulinum toxin (see AAX25521) was assembled from 3 separate
CC   toxin fragments using PCR and site-directed mutagenesis. The modified
CC   recombinant botulinum toxin maintains its ability to translocate from the
CC   gut into the general circulation but is non-toxic. It can be used as an
CC   oral vaccine for antigenic peptides including botulinum toxin (i.e. an
CC   oral vaccine for botulism) or for the oral delivery of other therapeutic
CC   agents to the general circulation
XX
SQ   Sequence 1291 AA;

  Query Match             33.4%;  Score 801.5;  DB 2;  Length 1291;
  Best Local Similarity   41.7%;  Pred. No. 1.2e-51;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||:||||||||||:|| |||||||||||    :|||| |:  |||| ||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 973

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        974 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| | |::  |         :|  :||||| ||:|||  :||||::
Db       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
                ||:|||||| |::: |||::| :|::||:   |  :|     : :    |  | ||
Db       1093 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 1152

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
              :      :|:  || :   |  |:: | :   :|:||     |   :||: |:   :
Db       1153 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 1209

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
                || |: : :   | |||| |:|  ||   :  |   :||      ::: |    |
Db       1210 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 1268

Qy        427 T--NYETKLLSTSSFWKFI 443
                || : | |||: | |:
Db       1269 KQGNYASLLESTSTHWGFV 1287


RESULT 7
AAW68394
ID   AAW68394 standard; protein; 472 AA.
XX
AC   AAW68394;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30581.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 303-305; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of recombinant proteins are E. coli, insect cells and yeast
CC   cells. The recombinant toxins are used as immunogens for the production
CC   of vaccines and antitoxins that are useful in the treatment of humans and
```

CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 32.2%; Score 772.5; DB 2; Length 472;
Best Local Similarity 37.3%; Pred. No. 4.7e-50;
Matches 186; Conservative 76; Mismatches 163; Indels 73; Gaps 14;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA--LLKDIINEYFNSINDSKILSLQNKKNALVDTS 58
              ||||||||||||||||||||||||||  :| :: |:| : | :: ||:|: | | |:| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYKDNNLIDLS 60

Qy         59 GYNAEVRVGDNVQLNTIYTNDFKLSSSGD-KIIVNLNNNILYSAIYENSSVSFWIKISKD 117
              || |:| | | |:||    | |||:|| : || |  | ||:::::: : ||||||:| |
Db         61 GYGAKVEVYDGVELND--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKY 118

Qy        118 LTNS-----HNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTG 172
                :.     ||||||| :: ||||||: ||   | | | |:| | ||: |:|:     :
Db        119 KNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178

Qy        173 YTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWI 232
              | |:|||||||||| :   |:||||:|: :  |:|: ||  :  |:| :| :||  | :|:
Db        179 YINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWM 237

Qy        233 RDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYI-----A 287
              : |:||: ||| :|   |: |     :||:||||| :: |||: |    : ||     :
Db        238 KYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDS 297

Qy        288 PESNVLVLVRYPDRSK------LYTGNPITIKSVSDKNPYSR--ILNGDNIILHMLYNSR 339
              |   :|   :|   ||      || |    |:  |:    :   :   | | |    ::
Db        298 PVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQ 357

Qy        340 KYMI-----------------IRDTDTIYAT-QGGECSQNCVYALKL--------QSNLG 373
              :: :                 | |:|  | | |  |  :  |: :|          :|
Db        358 EWRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIG 417

Qy        374 NYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKL 433
                 ||  |    ||          | |         || :    |      |    | ||
Db        418 LIGIHRFYESGIV------------FEE--------YKDYFCISKWYLKEVKRKPYNLKL 457

Qy        434 LSTSSFWKFISRDPGWVE 451
                    |:|| :| || |
Db        458 ---GCNWQFIPKDEGWTE 472
```

RESULT 8
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 31.9%; Score 765.5; DB 2; Length 472;
Best Local Similarity 36.4%; Pred. No. 1.6e-49;
Matches 183; Conservative 87; Mismatches 150; Indels 83; Gaps 17;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA--LLKDIINEYFNSINDSKILSLQNKKNALVDTS 58
              ||||||||||||||||||||||||||  :| :: |:| : | :: ||:|: | | |:| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIEMFNKYNSEILNNIILNLRYRDNNLIDLS 60

Qy         59 GYNAEVRVGDNVQLNTIYTNDFKLSSSGD-KIIVNLNNNILYSAIYENSSVSFWIKISK- 116
              || |:| | | |:||    | |||:|| | || |  | ||:::::: : ||||||:| |
Db         61 GYGAKVEVYDGVKLND--KNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKY 118

Qy        117 ---DLTN-SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTG 172
                  |: |  |||||||| :: |||||: ||    | | | |:| | ||: |:|:     :
Db        119 RNDDIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISE 178

Qy        173 YTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWI 232
              | |:|||||||||| :   |:|||| |: :  |:|: || ::  | | :| ::|  | :|:
Db        179 YINRWFFVTITNN-LDNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWM 237

Qy        233 RDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYI-----A 287
              : |:||: :|:  :|  :|: |     :||:||||| :: |||: |    : ||     :
Db        238 KYFSIFNTQLNQSNIKEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDS 297

Qy        288 PESNVLVLVRYPDRS------KLYTGNPITIKSVSDKNPYSRILNGD----------NII 331
                :|:  :|   |        || |    |: |:    |: :| |          :::
Db        298 SVGEILIRSKYNQNSNYINYRNLYIGEKFIIRRESN----SQSINDDIVRKEDYIHLDLV 353

Qy        332 LH----MLYNSRKYM----------IIRDTDTIYAT-QGGECSQNCVYALKL-------- 368
              ||      :| : ||           || |::  | | :  |   :    |: :|
Db        354 LHHEEWRVY-AYKYFKEQEEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEES 412

Qy        369 QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTN 428
              ::|  ||  |    :: |                    || :   |     |
Db        413 TDDIGLIGIHRFYESGVLRKK-------------------YKDYFCISKWYLKEVKRKP 452

Qy        429 YETKLLSTSSFWKFISRDPGWVE 451
              |:: |      |:|| :| || |
Db        453 YKSNL---GCNWQFIPKDEGWTE 472
```

RESULT 9
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX

```
AC   AAW68400;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW   botulism; BotG.
XX
OS   Clostridium botulinum; serotype G strain 113/30.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30596.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC   sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC   express BotG soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match             31.8%;  Score 762;  DB 2;  Length 473;
  Best Local Similarity   34.6%;  Pred. No. 3e-49;
  Matches  172;  Conservative   79;  Mismatches  176;  Indels   70;  Gaps   12;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA--LLKDIINEYFNSINDSKILSLQNKKNALVDTS 58
              |||||||||||||||||||||||||  :|  : | | ::|: : ||||  :   |:|:|
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         59 GYNAEVRVGDNVQLNTIYTNDFKLSSS-GDKIIVNLNNNILYSAIYENSSVSFWIKISKD 117
              || | : || :|  | |    |||::|    |  : :  ::| ::::| |::||::  |
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        118 LTNS-----HNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTG 172
                 |     ||||||: |: :||||: |:   | | | ||| | ||: |:||   : :
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        173 YTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWI 232
              | |||| :||||: :|   :|||| ||:|:|| :|| :     | | :    | : :||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        233 RDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYI------ 286
              :||||| :||:  ::: :|  |   | :||:|||||::||:||: |    : ||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        287 -----APESNV-LVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNII--------- 331
                  || :|     : |     || |    ||  |:    || :| |||:
Db        301 SMGETAPRTNFNNAAINY---QNLYLGLRFIIKKASN----SRNINNDNIVREGDYIYLN 353

Qy        332 ----------LHMLYNSRKYMI------IRDTDTIY-ATQGGECSQNCVYALKLQSNLGN 374
                       :::| ||::          | |  | |   |  :  :   |  ::
Db        354 IDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDT 413

Qy        375 YGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL 434
              |:| |   |    |   : ::              |    |    |    |
Db        414 KTFGLFGIGKFVKDYGYVWDTYDNY--------------FCISQWYLRRISENINKLRL 458

Qy        435 STSSFWKFISRDPGWVE 451
                :  |:||  | || |
Db        459 GCN--WQFIPVDEGWTE 473


RESULT 10
AAR95009
ID   AAR95009 standard; protein; 462 AA.
XX
AC   AAR95009;
XX
DT   07-JUL-1996  (first entry)
XX
DE   Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW   Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW   Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= Polyhistidine_tag
FT   Protein         22. .462
FT                   /label= C-fragment
XX
PN   WO9612802-A1.
XX
PD   02-MAY-1996.
XX
PF   23-OCT-1995;   95WO-US013737.
XX
PR   24-OCT-1994;   94US-00329154.
PR   16-MAR-1995;   95US-00405496.
PR   14-APR-1995;   95US-00422711.
PR   07-JUN-1995;   95US-00480604.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Padhye NV,  Kink JA,  Thalley BS,  Stafford DC;
PI   Firca JR;
XX
DR   WPI; 1996-230603/23.
DR   N-PSDB; AAT29246.
XX
PT   Fusion proteins comprising non-toxin protein and part of toxin - useful
PT   to form anti-toxins against Clostridium botulinum type A, and C.
PT   difficile type toxins, and to treat C. difficile intoxication, partic.
PT   diarrhoea.
XX
PS   Claim 7; Page 340-342; 434pp; English.
XX
CC   pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC   (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC   fragment C (see also AAR95008) of the Clostridium botulinum type A
CC   neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC   soluble protein and was purified by metal chelate affinity chromatography
```

CC to obtain a product free of endotoxin contamination that may be useful as
CC an immunogen in vaccine compsns
XX
SQ Sequence 462 AA;

Query Match 31.2%; Score 749; DB 2; Length 462;
Best Local Similarity 35.3%; Pred. No. 2.8e-48;
Matches 168; Conservative 92; Mismatches 174; Indels 42; Gaps 14;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||| |     ||  :| :: ||:|: : | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVRVGDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK- 116
                ::: :|  |  : |  |     | | ||  || | | | |:|:::||| | ||||:| |
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        117 -DLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTN 175
               :  : :|||||| :| |||||: :  | | | |||     : ::| ||: :: : | |
Db        119 FNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYIN 178

Qy        176 KWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDF 235
              :| ||||||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: |
Db        179 RWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYF 238

Qy        236 NIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID-------- 283
              |:| |||: ::|  :|: |    ::|||:||: |::|  ||::|    : |:|
Db        239 NLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG 298

Qy        284 -RYI-APESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKY 341
               |:  |  :|:    | : | || |    ||  :  |  : : | | : ::::  :::|
Db        299 YMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEY 357

Qy        342 MIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFR 400
              :  :     |:| |   :  : ||::  ::||   : :   ||        |
Db        358 RLATN-----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNN 409

Qy        401 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :     ::     | |    || :|:  : :  ||      |:||  | ||| |
Db        410 GNDIGFIGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459
```

RESULT 11
AAW68390
ID AAW68390 standard; protein; 462 AA.
XX
AC AAW68390;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism.
XX
OS Clostridium botulinum; serotype A.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal histidine tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30572, AAV30575.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 24; Page 265-267; 428pp; English.
XX
CC This is the amino acid sequence of a histidine-tagged fragment C
CC polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on a poly-histidine binding
CC affinity resin. The invention relates to recombinant proteins derived
CC from C. botulinum toxins. Methods are provided which allow the isolation
CC of soluble recombinant proteins that are free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant proteins are
CC E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are useful
CC in the treatment of humans and animals at risk of intoxication with
CC clostridial toxin
XX
SQ Sequence 462 AA;

Query Match 31.2%; Score 749; DB 2; Length 462;
Best Local Similarity 35.3%; Pred. No. 2.8e-48;
Matches 168; Conservative 92; Mismatches 174; Indels 42; Gaps 14;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||| |     ||  :| :: ||:|: : | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVRVGDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK- 116
                ::: :|  |  : |  |     | | ||  || | | | |:|:::||| | ||||:| |
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        117 -DLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTN 175
               :  : :|||||| :| |||||: :  | |. | |||     : ::| ||: :: : | |
Db        119 FNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYIN 178

Qy        176 KWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDF 235
              :| ||||||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: |
Db        179 RWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYF 238

Qy        236 NIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID-------- 283
              |:| |||: ::|  :|: |    ::|||:||: |::|  ||::|    : |:|
Db        239 NLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG 298

Qy        284 -RYI-APESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKY 341
               |:  |  :|:    | : | || |    ||  :  |  : : | | : ::::  :::|
Db        299 YMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEY 357

Qy        342 MIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFR 400
              :  :     |:| |   :  : ||::  ::||   : :   ||        |
Db        358 RLATN-----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNN 409

Qy        401 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :     ::     | |    || :|:  : :  ||      |:||  | ||| |
Db        410 GNDIGFIGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459
```

RESULT 12
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.

```
XX
KW   Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW   botulism; BotF.
XX
OS   Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 48; Page 364-365; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC   sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC   express BotC soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 448 AA;

  Query Match             29.1%;  Score 698.5;  DB 2;  Length 448;
  Best Local Similarity   33.7%;  Pred. No. 1.8e-44;
  Matches  166;  Conservative   80;  Mismatches  160;  Indels   87;  Gaps   15;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||:|    |  :  | || || :: : |  :| |||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSG-DKIIVNLNNNILYSAIYENSSVSFWIKISKDL- 118
                : : :  || : :   | | : :|   :: :  ||:|:|:: |:| |:|||::| |
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        119 -TNSHNEYTIINSI-EQNSGWKL---CIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGY 173
               | : |||||| :   |||||:    :|:  | | ||| :   ::||| | |    : |
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        174 TNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIR 233
               ||| ||||||| :|  ::|||| |   : | :| :: :   |:| |    |:   : ||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKI-VGCDDETYVGIR 239

Qy        234 DFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVL 293
               | :|: ||   :|  :|  :   :::|:|||| | :: :||: |    |:||   | :|
Db        240 YFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGIL 299

Qy        294 ------------VLVRYPDRSKLYTG--------NPITIKSVSD---KNPYSRILNGDNI 330
                                | : |    ||| |          || | :  :   ||  : |   |
Db        300 NINQQRGVTEGSVFLNY----KLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRG 355

Qy        331 ILHMLY---NSRKYMIIRDTD--------TIYATQGGECSQNCVYALKLQSNLG-NYGIG 378
              : : ||    | |  ||| ::        :  : |  |: |       |:| | | |:
Db        356 VEYRLYADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMN------FQNNNGSNIGLL 409

Qy        379 IFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS 438
               |   |:|: : |    ::: | ||                                |
Db        410 GFHSNNLVASSWY----YNNIRRNTS----------------------------SNGC 435

Qy        439 FWKFISRDPGWVE 451
              ||  ||:: || |
Db        436 FWSSISKENGWKE 448


RESULT 13
AAW68396
ID   AAW68396 standard; protein; 452 AA.
XX
AC   AAW68396;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain NCTC 11219.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;
```

```
  Query Match             28.9%;  Score 693;  DB 2;  Length 452;
  Best Local Similarity   35.1%;  Pred. No. 4.7e-44;
  Matches  170;  Conservative   83;  Mismatches  164;  Indels   68;  Gaps   20;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAL--------LKDIINEYFNSINDSKILSLQNKKN 52
              ||||||||||||||||||||||||||||        |    |::|  |  | :|::: | :
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 60

Qy         53 ALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSGDKII-VNLNNN--ILYSAIYENSSVS 109
                ||||||:: : :  :|       | | : :  ||:  ||:: |  |:|   |:| |:|
Db         61 KYVDTSGYDSNININGDVYKYPTNKNQFGIDN--DKLSEVNISQNDYIIYDNKYKNFSIS 118

Qy        110 FWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQD---VNRKYKSLIF 162
              ||::|      : | :|||||||| : : |||||: : :  | | |||   :|:|   | |
Db        119 FWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQK---LAF 175

Qy        163 DYSESLSHTGYTNKWPFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDE 222
              :|  :   : | ||| ||||||: :|  |||||| |   : | :|  : :   |:| | |
Db        176 NYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKI-V 234

Qy        223 NIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN---- 278
              |   : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|
Db        235 NCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKP 294

Qy        279 DNYIDRYIAP-------ESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNII 331
              :|:|||             | :|:       ::||:|  : |: |:: :         ||::
Db        295 NNFIDRRKDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSS------TNDNLV 342

Qy        332 LHMLYNSRKYM-IIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNK 390
                  | : |:  :     ::       : |    :|: |:   :  :  : | |  |
Db        343 RK---NDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVV--VMNSVGNN- 396

Qy        391 YCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFISRD 446
              |:  | :   | : |             || | | || | | | :     |   || ||| :
Db        397 -CTMNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDHTNSNGCFWNFISEE 446

Qy        447 PGWVE 451
              || |
Db        447 HGWQE 451


RESULT 14
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             28.6%;  Score 684.5;  DB 2;  Length 451;
  Best Local Similarity   34.2%;  Pred. No. 2.1e-43;
  Matches  166;  Conservative   85;  Mismatches  165;  Indels   69;  Gaps   15;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAL--------LKDIINEYFNSINDSKILSLQNKKN 52
              ||||||||||||||||||||||||||||        |    |::|  |  | :|::: | :
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 60

Qy         53 ALVDTSGYNAEVRVGDNVQLNTIYTNDF-----KLSSSGDKIIVNLNNNILYSAIYENSS 107
                ||||||:: : :  :|       | |     |||    :: :: |: |:|    |:| |
Db         61 KYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLS----EVNISQNDYIIYDNKYKNFS 116

Qy        108 VSFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEWILQDVNRKYKSLIFD 163
              :|||::|      : | :|||||||| : : |||||: : :  | | ||| :   : | |:
Db        117 ISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFN 176

Qy        164 YSESLSHTGYTNKWPFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDEN 223
              |  :   : | ||| ||||||: :|  |||||| |   : | :|  : :   |:| | |
Db        177 YGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKI-VN 235

Qy        224 IDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----D 279
              : : || |||| |||   :|  :|  :   |::||:||| | :| |||::|    :
Db        236 CSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPN 295

Qy        280 NYIDRYIAP-------ESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIIL 332
              |:|:|             | :|:       ::||:|  : |: |:: :    ::  :: :
Db        296 NFINRRTDSTLSINNIRSTILL------ANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVY 349

Qy        333 HMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQS--NLGNYGIGIFSIKNIVSKNK 390
                  |: :::    ||          : |    :|: |  |  |  : : |: |
Db        350 INFVASKTHLLPLYADT--------ATTNKEKTIKISSSGNRFNQVVVMNSVGN------ 395

Qy        391 YCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLL----STSSFWKFISRD 446
              |:  | :   | : |             || | | || | | | :     |   || ||| :
Db        396 -CTMNFKNNNGNNIGL--------LGFK-ADTVVASTWYYTHMRDNTNSNGFFWNFISEE 445

Qy        447 PGWVE 451
              || |
Db        446 HGWQE 450


RESULT 15
AAW68391
ID   AAW68391 standard; protein; 445 AA.
XX
AC   AAW68391;
XX
DT   07-DEC-1998  (first entry)
```

XX
DE Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism.
XX
OS Clostridium botulinum; serotype A.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .7
FT /note= "N-terminal histidine tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30576.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 29; Page 279-281; 428pp; English.
XX
CC This is the amino acid sequence of a histidine-tagged fragment C
CC polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC sequence (see AAV30576) in plasmid p6HisBot(syn). This vector was used to
CC express native soluble C fragment in Escherichia coli host cells, with
CC the recombinant C fragment being purified on a poly-histidine binding
CC affinity resin. The invention relates to recombinant proteins derived
CC from C. botulinum toxins. Methods are provided which allow the isolation
CC of soluble recombinant proteins that are free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant proteins are
CC E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are useful
CC in the treatment of humans and animals at risk of intoxication with
CC clostridial toxin
XX
SQ Sequence 445 AA;

Query Match 26.2%; Score 627; DB 2; Length 445;
Best Local Similarity 31.9%; Pred. No. 4.6e-39;
Matches 150; Conservative 92; Mismatches 174; Indels 54; Gaps 15;

```
Qy          7 HHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRV 66
              ||||||            || |     ||  :| :: ||:|: : | |:| | | ::: :
Db          2 HHHHHH-----------MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 49

Qy         67 GDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNS 121
                |  | : |  |    | | ||  || | | | |:|:::||| | ||||:| |  :  :
Db         50 GSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISL 107

Qy        122 HNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVT 181
              :||||||| :| ||||| :  | | | |||     : ::| ||: :: : | |:| |||
Db        108 NNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVT 167

Qy        182 ITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKE 241
              |||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: ||:| ||
Db        168 ITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKE 227

Qy        242 LSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID---------RYI-A 287
              |: ::|  :|: |    ::||:||: |::|  ||::|    : |:|          |:
Db        228 LNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKG 287

Qy        288 PESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDT 347
              |  :|:    | : | || |    ||  :  |  : : | | : ::::  :::| :  :
Db        288 PRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN- 345

Qy        348 DTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLL 406
                |:| |   :  : ||::  ::||   : :   ||       |          | :
Db        346 ----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGF 398

Qy        407 ADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
                ::    | |    || :|:  : :  ||       |:||  | || |
Db        399 IGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 442
```

Search completed: September 10, 2004, 03:33:41
Job time : 31.7495 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 9.57257 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 2397
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	749	31.2	462	1	US-08-480-604A-26	Sequence 26, Appl
2	749	31.2	462	2	US-08-405-496A-26	Sequence 26, Appl
3	749	31.2	462	3	US-08-915-136-26	Sequence 26, Appl
4	749	31.2	462	4	US-09-084-517-26	Sequence 26, Appl
5	596	24.9	1169	4	US-09-255-829-20	Sequence 20, Appl
6	595	24.8	438	1	US-08-480-604A-23	Sequence 23, Appl
7	595	24.8	438	2	US-08-405-496A-23	Sequence 23, Appl
8	595	24.8	438	3	US-08-915-136-23	Sequence 23, Appl
9	595	24.8	438	4	US-09-084-517-23	Sequence 23, Appl
10	588	24.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
11	588	24.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
12	588	24.5	1296	3	US-08-915-136-28	Sequence 28, Appl
13	588	24.5	1296	4	US-09-084-517-28	Sequence 28, Appl
14	401.5	16.8	452	1	US-07-618-312A-2	Sequence 2, Appli
15	401.5	16.8	452	1	US-07-618-312A-4	Sequence 4, Appli
16	401.5	16.8	452	1	US-08-280-228-2	Sequence 2, Appli
17	401.5	16.8	452	1	US-08-280-228-4	Sequence 4, Appli
18	401.5	16.8	618	1	US-08-668-381A-5	Sequence 5, Appli
19	401.5	16.8	853	4	US-08-913-880C-17	Sequence 17, Appl
20	401.5	16.8	858	4	US-08-913-880C-16	Sequence 16, Appl
21	401.5	16.8	860	4	US-08-913-880C-15	Sequence 15, Appl
22	401.5	16.8	862	4	US-08-913-880C-14	Sequence 14, Appl
23	401.5	16.8	865	4	US-08-913-880C-13	Sequence 13, Appl
24	401.5	16.8	866	4	US-08-913-880C-12	Sequence 12, Appl
25	401.5	16.8	874	4	US-08-913-880C-11	Sequence 11, Appl
26	401.5	16.8	875	4	US-08-913-880C-10	Sequence 10, Appl
27	401.5	16.8	1315	4	US-08-913-880C-1	Sequence 1, Appli
28	370.5	15.5	452	1	US-08-110-786A-8	Sequence 8, Appli
29	153	6.4	1534	4	US-09-543-681A-5182	Sequence 5182, Ap
30	150	6.3	23	1	US-08-480-604A-24	Sequence 24, Appl
31	150	6.3	23	2	US-08-405-496A-24	Sequence 24, Appl
32	150	6.3	23	3	US-08-915-136-24	Sequence 24, Appl
33	150	6.3	23	4	US-09-084-517-24	Sequence 24, Appl
34	146	6.1	22	4	US-09-660-742-3	Sequence 3, Appli
35	146	6.1	53	2	US-08-651-818A-19	Sequence 19, Appl
36	146	6.1	53	3	US-09-184-826-19	Sequence 19, Appl
37	146	6.1	54	2	US-08-651-818A-23	Sequence 23, Appl
38	146	6.1	54	3	US-09-184-826-23	Sequence 23, Appl
39	146	6.1	323	1	US-08-591-989-5	Sequence 5, Appli
40	141	5.9	21	2	US-08-651-818A-21	Sequence 21, Appl
41	141	5.9	21	3	US-09-184-826-21	Sequence 21, Appl
42	140	5.8	476	3	US-09-316-083-3	Sequence 3, Appli
43	140	5.8	476	4	US-09-933-700-3	Sequence 3, Appli
44	134.5	5.6	945	4	US-09-543-681A-4200	Sequence 4200, Ap
45	131.5	5.5	2391	2	US-08-446-855A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-26

Query Match 31.2%; Score 749; DB 1; Length 462;
Best Local Similarity 35.3%; Pred. No. 4.7e-53;
Matches 168; Conservative 92; Mismatches 174; Indels 42; Gaps 14;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||| |     ||  :| :: ||:|: : | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVRVGDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK- 116
               ::: :|  |  : |  |    | | ||  || | | |:|:::|||  | ||||:| |
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        117 -DLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTN 175
              :  : :|||||| :| |||||: :  | | | |||     : ::| ||: :: : | |
Db        119 FNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYIN 178

Qy        176 KWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDF 235
              :| |||||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: |
Db        179 RWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYF 238

Qy        236 NIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID-------- 283
              |:| |||: ::|  :|: |    ::||:||: |::|  ||::|    : |:|
Db        239 NLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG 298

Qy        284 -RYI-APESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKY 341
               |:  |  :|:   | : | || |   ||  :  |  : : | | : ::::  :::|
Db        299 YMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEY 357

Qy        342 MIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFR 400
              :  :     |:| |   :  : ||::  ::||   : :   ||       |
Db        358 RLATN-----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNN 409

Qy        401 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :    ::     | |    || :|:  : :  ||      |:||  | || |
Db        410 GNDIGFIGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459
```

RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-26

Query Match 31.2%; Score 749; DB 2; Length 462;
Best Local Similarity 35.3%; Pred. No. 4.7e-53;
Matches 168; Conservative 92; Mismatches 174; Indels 42; Gaps 14;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||| |     ||  :| :: ||:|: : | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVRVGDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK- 116
               ::: :|  |  : |  |    | | ||  || | | |:|:::|||  | ||||:| |
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        117 -DLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTN 175
              :  : :|||||| :| |||||: :  | | | |||     : ::| ||: :: : | |
Db        119 FNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYIN 178

Qy        176 KWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDF 235
              :| |||||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: |
Db        179 RWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYF 238

Qy        236 NIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID-------- 283
              |:| |||: ::|  :|: |    ::||:||: |::|  ||::|    : |:|
Db        239 NLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG 298

Qy        284 -RYI-APESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKY 341
               |:  |  :|:   | : | || |   ||  :  |  : : | | : ::::  :::|
Db        299 YMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEY 357

Qy        342 MIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFR 400
              :  :     |:| |   :  : ||::  ::||   : :   ||       |
Db        358 RLATN-----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNN 409

Qy        401 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :    ::     | |    || :|:  : :  ||      |:||  | || |
Db        410 GNDIGFIGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459
```

RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.

```
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match           31.2%;  Score 749;  DB 3;  Length 462;
  Best Local Similarity  35.3%;  Pred. No. 4.7e-53;
  Matches  168;  Conservative   92;  Mismatches  174;  Indels   42;  Gaps   14;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||| |     ||  :| :: ||:|: : | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVRVGDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK- 116
               ::: :|  |  : |  |    | | ||  || | | |:|:::||| | ||||:| |
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        117 -DLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTN 175
              :  : :||||||| :| ||||||: :  | | | |||     : ::| ||: :: : | |
Db        119 FNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYIN 178

Qy        176 KWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDF 235
              :| ||||||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: |
Db        179 RWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYF 238

Qy        236 NIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID-------- 283
              |:| |||: ::|  :|: |    ::||:||: |::|  ||::|    : |:|
Db        239 NLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG 298

Qy        284 -RYI-APESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKY 341
               |:  |  :|:    | : | || |    ||  :  |  : : | | : :::: :::|
Db        299 YMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEY 357

Qy        342 MIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFR 400
              :  :     |:| |   :  : ||::  ::||  : :   ||       |
Db        358 RLATN-----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNN 409

Qy        401 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :    ::    | |    || :|:  : :  ||      |:||  | || |
Db        410 GNDIGFIGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26
```

```
  Query Match             31.2%;  Score 749;  DB 4;  Length 462;
  Best Local Similarity   35.3%;  Pred. No. 4.7e-53;
  Matches  168;  Conservative   92;  Mismatches  174;  Indels   42;  Gaps   14;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||| |      ||  :| :: ||:|: : | |:| | |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 NAEVRVGDNVQLNTIYTND---FKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK- 116
               ::: :|  |  : |  |    | | ||  || | | |:|:::||| | ||||:| |
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        117 -DLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTN 175
              :  : :|||||| :| |||||: :  | | | |||     : ::| ||: :: : | |
Db        119 FNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYIN 178

Qy        176 KWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDF 235
              :| ||||||| :   |:|||| |   : | :|  :     |:| :|   | :: :||: |
Db        179 RWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYF 238

Qy        236 NIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIIN----DNYID-------- 283
              |:| |||: ::|  :|: |    ::||:||: |::|  ||::|    : |:|
Db        239 NLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG 298

Qy        284 -RYI-APESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKY 341
               |:  |  :|:    | : | || |    ||  :  |  : : | | : :::  :::|
Db        299 YMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEY 357

Qy        342 MIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFR 400
              :  :     |:| |   :  : ||::  ::||   : :   ||        |
Db        358 RLATN-----ASQAG--VEKILSALEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNN 409

Qy        401 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :     ::    | |    || :|:  : :  ||       |:||  | || |
Db        410 GNDIGFIGFHQ-----FNNIAKLVA-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459


RESULT 5
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-20

  Query Match             24.9%;  Score 596;  DB 4;  Length 1169;
  Best Local Similarity   39.2%;  Pred. No. 5.1e-40;
  Matches  134;  Conservative   64;  Mismatches  106;  Indels   38;  Gaps    8;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              :| :: |:| : | :: ||:|: | | |:| ||| |:| | | |:||    | |||:||
Db        848 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA 905

Qy         87 D-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
              : || |  | ||:::::: : ||||||:| |    :       |||||||| :: |||||:
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              ||   | | | |:| | ||: |:|:     : | |:|||||||| :   |:||||:|:
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLES 1024

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
              :  |:|: ||  :  |:| :| :||  | :|:: |:||: |||  :|   |: |     :
Db       1025 NTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYL 1084

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRSK------LYTGNP 309
              ||:||||| :: |||: |    : ||     :|   :|   :|   ||      || |
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 1144

Qy        310 ITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIY 351
                |:  |:    |: :| |               |:|  | ||
Db       1145 FIIRRKSN----SQSINDD--------------IVRKEDYIY 1168


RESULT 6
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match             24.8%;  Score 595;  DB 1;  Length 438;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-40;
  Matches  144;  Conservative   92;  Mismatches  174;  Indels   42;  Gaps   14;

Qy         25 MALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FK 81
              || |     ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKL 139
              | ||  || | | | |:|:::||| | ||||:| |  :  : :||||||| :| |||||:
Db         61 LESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 118

Qy        140 CIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELK 199
               : .| | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        200 QSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNV 259
               : | :|  :     |:| :|    | :: :||: ||:| |||: ::|  :|: |    :
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        260 IKDYWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLY 305
              :||:||: |::|  ||::|    : |:|          |:  |  :|:    | : | ||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        306 TGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYA 365
              |    ||  :  |  : : | | : ::::  :::| :  :     |:| |   :  : |
Db        298 RGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSA 350

Qy        366 LKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPV 424
              |::  ::||   : :  ||       |         | :    ::     | |    |
Db        351 LEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLV 404

Qy        425 AVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :|:  : :  ||      |:||  | || |
Db        405 A-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 435


RESULT 7
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match             24.8%;  Score 595;  DB 2;  Length 438;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-40;
  Matches  144;  Conservative   92;  Mismatches  174;  Indels   42;  Gaps   14;

Qy         25 MALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FK 81
              || |     ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKL 139
              | ||  || | | | |:|:::||| | ||||:| |  :  : :|||||| :| |||||:
Db         61 LESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 118

Qy        140 CIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELK 199
               : | | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        200 QSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNV 259
               : | :|  :     |:| :|    | :: :||: ||:| |||: ::|  :|: |    :
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        260 IKDYWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLY 305
              :||:||: |::|  ||::|    : |:|          |:  |  :|:    | : | ||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        306 TGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYA 365
              |    ||  :  |  : : | | : ::::  :::| :  :     |:| |   :  : |
Db        298 RGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSA 350
```

```
Qy        366 LKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPV 424
              |::  ::||   : :   ||       |          | :     ::      | |    |
Db        351 LEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLV 404

Qy        425 AVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :|:  : :  ||      |:||  | || |
Db        405 A-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 435


RESULT 8
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match             24.8%;  Score 595;  DB 3;  Length 438;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-40;
  Matches  144;  Conservative   92;  Mismatches  174;  Indels   42;  Gaps   14;

Qy         25 MALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FK 81
              ||  |      ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |     |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKL 139
              | ||  || | | | |:|:::||| | ||||:| |  :  : :|||||| :| |||||:
Db         61 LESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 118

Qy        140 CIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELK 199
              :  | | | |||      : ::| ||: :: : | |:| ||||||| :   |:|||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        200 QSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNV 259
              : | :|  :      |:| :|   | :: :||: ||:| |||: ::|  :|: |     :
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        260 IKDYWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLY 305
              :||:||: |::|  ||::|    : |:|           |:  |  :|:    | : | ||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        306 TGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYA 365
              |    ||  :  |  : : | | : ::::  :::| :  :      |:| |   :  : |
Db        298 RGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSA 350

Qy        366 LKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPV 424
              |::  ::||   : :   ||       |          | :     ::      | |    |
Db        351 LEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLV 404

Qy        425 AVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :|:  : :  ||      |:||  | || |
Db        405 A-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 435


RESULT 9
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-23

  Query Match             24.8%;  Score 595;  DB 4;  Length 438;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-40;
  Matches  144;  Conservative   92;  Mismatches  174;  Indels   42;  Gaps   14;

Qy         25 MALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FK 81
              |  |      ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    |
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKL 139
              | ||  || | | | |:|:::||| | ||||:| |  :  : :||||||| :| |||||:
Db         61 LESS--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV 118

Qy        140 CIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELK 199
               :  | | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |
Db        119 SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLI 178

Qy        200 QSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNV 259
               : | :|  :     |:| :|   | :: :||: ||:| |||: ::|  :|: |    :
Db        179 DQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGI 238

Qy        260 IKDYWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLY 305
              :||:||: |::|  ||::|    : |:|         |:  |  :|:    | : | ||
Db        239 LKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLY 297

Qy        306 TGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYA 365
              |    ||  :  |  : : | | : :::  :::| :  :     |:| |   :  : |
Db        298 RGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSA 350

Qy        366 LKLQSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPV 424
              |::  ::||   : :   ||        |          | :     ::      | |   |
Db        351 LEI-PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLV 404

Qy        425 AVTNYETKLLSTSS-----FWKFISRDPGWVE 451
              | :|:  : :  ||      |:||  | || |
Db        405 A-SNWYNRQIERSSRTLGCSWEFIPVDDGWGE 435


RESULT 10
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             24.5%;  Score 588;  DB 1;  Length 1296;
  Best Local Similarity   31.6%;  Pred. No. 2.6e-39;
  Matches  142;  Conservative   92;  Mismatches  173;  Indels   42;  Gaps   14;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FKLSS 84
              |      ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKLCIR 142
              |  || | | | |:|:::||| | ||||:| |  :  : :||||||| :| |||||: :
Db        922 S--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLN 979

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
               | | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |   :
Db        980 YGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQK 1039

Qy        203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
               | :|  :     |:| :|    | :: :||: ||:| |||: ::|  :|: |    ::||
Db       1040 PISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKD 1099

Qy        263 YWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGN 308
              :||: |::|  ||::|    : |:|         |:  |  :|:    | : | || |
Db       1100 FWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGT 1158

Qy        309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
              ||  :  |  : : | | : ::::  :::| :  :     |:| |   :  : ||::
Db       1159 KFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSALEI 1211

Qy        369 QSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVT 427
```

```
                ::||   : :   ||       |         | :     ::      | |    || :
Db       1212 -PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLVA-S 1264

Qy        428 NYETKLLSTSS-----FWKFISRDPGWVE 451
              |:  : :  ||      |:||  | || |
Db       1265 NWYNRQIERSSRTLGCSWEFIPVDDGWGE 1293


RESULT 11
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             24.5%;  Score 588;  DB 2;  Length 1296;
  Best Local Similarity   31.6%;  Pred. No. 2.6e-39;
  Matches  142;  Conservative   92;  Mismatches  173;  Indels   42;  Gaps   14;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FKLSS 84
              |     ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKLCIR 142
              |  || | | | |:|:::||| | ||||:| |  :  : :|||||| :| |||||: :
Db        922 S--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLN 979

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
                | | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |   :
Db        980 YGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQK 1039

Qy        203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
              | :|  :     |:| :|    | :: :||: ||:| |||: ::|  :|: |    ::||
Db       1040 PISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKD 1099

Qy        263 YWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGN 308
              :||: |::|  ||::|    : |:|          |:  |  :|:    | : | || |
Db       1100 FWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGT 1158

Qy        309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
              ||  :  |  : : | | : ::::  :::| :  :       |:| |   :  : ||::
Db       1159 KFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSALEI 1211

Qy        369 QSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVT 427
              ::||   : :   ||       |         | :     ::     | |    || :
Db       1212 -PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLVA-S 1264

Qy        428 NYETKLLSTSS-----FWKFISRDPGWVE 451
              |:  : :  ||      |:||  | || |
Db       1265 NWYNRQIERSSRTLGCSWEFIPVDDGWGE 1293


RESULT 12
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
```

```
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             24.5%;  Score 588;  DB 3;  Length 1296;
  Best Local Similarity   31.6%;  Pred. No. 2.6e-39;
  Matches  142;  Conservative   92;  Mismatches  173;  Indels   42;  Gaps   14;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FKLSS 84
              |      ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |     | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKLCIR 142
              |  || | | | |:|:::||| | ||||:| |  :  : :||||||| :| |||||: :
Db        922 S--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLN 979

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              | | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |   :
Db        980 YGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQK 1039

Qy        203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
              | :|  :     |:| :|   | :: :||: ||:| |||: ::|  :|: |    ::||
Db       1040 PISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKD 1099

Qy        263 YWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGN 308
              :||: |::|  ||::|    : |:|         |:  |  :|:    | : | || |
Db       1100 FWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGT 1158

Qy        309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
              ||  :  |  : : | | : ::::  :::| :  :     |:| |   :  : ||::
Db       1159 KFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSALEI 1211

Qy        369 QSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVT 427
               ::||   : :   ||        |          | :     ::     | |    || :
Db       1212 -PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLVA-S 1264

Qy        428 NYETKLLSTSS-----FWKFISRDPGWVE 451
              |:  : :  ||      |:||  | || |
Db       1265 NWYNRQIERSSRTLGCSWEFIPVDDGWGE 1293


RESULT 13
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             24.5%;  Score 588;  DB 4;  Length 1296;
  Best Local Similarity   31.6%;  Pred. No. 2.6e-39;
  Matches  142;  Conservative   92;  Mismatches  173;  Indels   42;  Gaps   14;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FKLSS 84
              |      ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |     | | |
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         85 SGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKLCIR 142
              |  || | | | |:|:::||| | ||||:| |  :  : :||||||| :| |||||: :
Db        922 S--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLN 979

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              | | | |||     : ::| ||: :: : | |:| ||||||| :   |:|||| |   :
Db        980 YGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQK 1039

Qy        203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
              | :|  :     |:| :|   | :: :||: ||:| |||: ::|  :|: |    ::||
Db       1040 PISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKD 1099

Qy        263 YWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGN 308
              :||: |::|  ||::|    : |:|         |:  |  :|:    | : | || |
Db       1100 FWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGT 1158

Qy        309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
              ||  :  |  : : | | : ::::  :::| :  :     |:| |   :  : ||::
Db       1159 KFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSALEI 1211

Qy        369 QSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVT 427
               ::||   : :   ||        |          | :     ::     | |    || :
Db       1212 -PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLVA-S 1264

Qy        428 NYETKLLSTSS-----FWKFISRDPGWVE 451
              |:  : :  ||      |:||  | || |
Db       1265 NWYNRQIERSSRTLGCSWEFIPVDDGWGE 1293


RESULT 14
US-07-618-312A-2
; Sequence 2, Application US/07618312A
```

```
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             16.8%;  Score 401.5;  DB 1;  Length 452;
  Best Local Similarity   24.8%;  Pred. No. 1e-24;
  Matches  118;  Conservative   92;  Mismatches  158;  Indels  107;  Gaps   16;

Qy         39 INDSKILSLQNKKNALVDTSGYNAEVRVGDNVQL------NTIYTNDFKLSSSGDKIIVN 92
              :  | ||:|    :  :| ||:|: |    : ||        |:     :::   ::||:
Db         18 LKKSTILNLDINNDINIDISGFNSSVITYPDAQLVPGINGKAIHL----VNNESSEVIVH 73

Qy         93 LNNNILYSAIYENSSVSFWIKISKDLTNSH------NEYTIINSIEQN-----SGWKLCI 141
                 :| |: :: | :||||::: | :: ||       |||:||:|::::     ||| : :
Db         74 KAMDIEYNDMFNNFTVSFWLRVPK-VSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSL 132

Qy        142 RNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQS 201
              :  |: | |:|   : : : |       :    ||| |:||||: :    |||||  |  |
Db        133 KGNNLIWTLKDSAGEVRQITFRDLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGS 192

Qy        202 QKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIK 261
               :|  |  :: |  |   :|   : || : |  | || | |: ::|  :|   :    ::
Db        193 AEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLR 252

Qy        262 DYWGNPLKFDTEYYII--NDNYIDRYIAPESNVLVLVRYPDRS---------KLYTGNPI 310
              |:|||||::|||||:|    :  |  :   :: : |   |  :          :|| |
Db        253 DFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKF 312

Qy        311 TIKSVSDKNPY-SRILNGDNIILHMLYNSRKYMI------------------------- 343
               ||  :  |   | : :|| | |:: ||: ::::
Db        313 IIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIP 372

Qy        344 ---------IRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQ 394
                       :||  | |: |          |||  :  |  :|:    |         |
Db        373 LYKKMEAVKLRDLKT-YSVQ----------LKLYDD-KNASLGLVGTHN--------GQ 411

Qy        395 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 449
              | :    : :: ::    | |:           : : |:|     | |:  | ||
Db        412 IGNDPNRDILIASN----WYFN----------HLKDKILGCD--WYFVPTDEGW 449


RESULT 15
US-07-618-312A-4
; Sequence 4, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-4

  Query Match             16.8%;  Score 401.5;  DB 1;  Length 452;
  Best Local Similarity   24.8%;  Pred. No. 1e-24;
  Matches  118;  Conservative   92;  Mismatches  158;  Indels  107;  Gaps   16;

Qy         39 INDSKILSLQNKKNALVDTSGYNAEVRVGDNVQL------NTIYTNDFKLSSSGDKIIVN 92
              :  | ||:|    : : | ||:|: |    : ||        |:     :::   ::||:
Db         18 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHL----VNNESSEVIVH 73

Qy         93 LNNNILYSAIYENSSVSFWIKISKDLTNSH------NEYTIINSIEQN-----SGWKLCI 141
                 :| |: :: | :||||::: | :: ||       |||:||:|::::     ||| : :
Db         74 KAMDIEYNDMFNNFTVSFWLRVPK-VSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSL 132

Qy        142 RNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQS 201
              :  |: | |:|   : : : |       :    ||| |:||||: :    |||||  |  |
```

```
Db        133 KGNNLIWTLKDSAGEVRQITFRDLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGS 192

Qy        202 QKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIK 261

Db        193 AEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLR 252

Qy        262 DYWGNPLKFDTEYYII--NDNYIDRYIAPESNVLVLVRYPDRS---------KLYTGNPI 310

Db        253 DFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKF 312

Qy        311 TIKSVSDKNPY-SRILNGDNIILHMLYNSRKYMI------------------------- 343

Db        313 IIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIP 372

Qy        344 ---------IRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQ 394

Db        373 LYKKMEAVKLRDLKT-YSVQ-----------LKLYDD-KNASLGLVGTHN--------GQ 411

Qy        395 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 449

Db        412 IGNDPNRDILIASN----WYFN-----------HLKDKILGCD--WYFVPTDEGW 449
```

Search completed: September 10, 2004, 03:47:10
Job time : 10.5726 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 87.2996 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 2397
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2397	100.0	451	12	US-10-272-898-68	Sequence 68, Appl
2	2397	100.0	451	15	US-10-354-774-68	Sequence 68, Appl
3	2397	100.0	451	15	US-10-271-012-68	Sequence 68, Appl
4	2397	100.0	451	16	US-10-729-122-68	Sequence 68, Appl
5	2397	100.0	451	16	US-10-729-039-68	Sequence 68, Appl
6	2235	93.2	1276	12	US-10-272-898-66	Sequence 66, Appl
7	2235	93.2	1276	15	US-10-354-774-66	Sequence 66, Appl
8	2235	93.2	1276	15	US-10-271-012-66	Sequence 66, Appl
9	2235	93.2	1276	15	US-10-452-024-164	Sequence 164, App
10	2235	93.2	1276	16	US-10-729-122-66	Sequence 66, Appl
11	2235	93.2	1276	16	US-10-729-039-66	Sequence 66, Appl
12	2231	93.1	1276	15	US-10-452-024-4	Sequence 4, Appli
13	2231	93.1	1276	15	US-10-452-024-168	Sequence 168, App
14	2231	93.1	1276	15	US-10-205-516-8	Sequence 8, Appli
15	2231	93.1	1286	15	US-10-205-516-22	Sequence 22, Appl
16	2226	92.9	1275	15	US-10-452-024-167	Sequence 167, App
17	2224	92.8	451	10	US-09-910-186A-12	Sequence 12, Appl
18	2131	88.9	1280	15	US-10-452-024-139	Sequence 139, App
19	2131	88.9	1280	15	US-10-452-024-140	Sequence 140, App
20	963.5	40.2	462	12	US-10-272-898-62	Sequence 62, Appl
21	963.5	40.2	462	15	US-10-354-774-62	Sequence 62, Appl
22	963.5	40.2	462	15	US-10-271-012-62	Sequence 62, Appl
23	963.5	40.2	462	16	US-10-729-122-62	Sequence 62, Appl
24	963.5	40.2	462	16	US-10-729-039-62	Sequence 62, Appl
25	801.5	33.4	450	10	US-09-910-186A-10	Sequence 10, Appl
26	801.5	33.4	1291	12	US-10-272-898-60	Sequence 60, Appl
27	801.5	33.4	1291	15	US-10-354-774-60	Sequence 60, Appl
28	801.5	33.4	1291	15	US-10-271-012-60	Sequence 60, Appl
29	801.5	33.4	1291	15	US-10-452-024-3	Sequence 3, Appli
30	801.5	33.4	1291	15	US-10-452-024-135	Sequence 135, App
31	801.5	33.4	1291	15	US-10-452-024-136	Sequence 136, App
32	801.5	33.4	1291	15	US-10-205-516-6	Sequence 6, Appli
33	801.5	33.4	1291	16	US-10-729-122-60	Sequence 60, Appl
34	801.5	33.4	1291	16	US-10-729-039-60	Sequence 60, Appl
35	801.5	33.4	1301	15	US-10-205-516-20	Sequence 20, Appl
36	782.5	32.6	1285	15	US-10-452-024-166	Sequence 166, App
37	781.5	32.6	1285	15	US-10-452-024-165	Sequence 165, App
38	772.5	32.2	472	12	US-10-272-898-46	Sequence 46, Appl
39	772.5	32.2	472	15	US-10-354-774-46	Sequence 46, Appl
40	772.5	32.2	472	15	US-10-271-012-46	Sequence 46, Appl
41	772.5	32.2	472	16	US-10-729-122-46	Sequence 46, Appl
42	772.5	32.2	472	16	US-10-729-039-46	Sequence 46, Appl
43	765.5	31.9	472	12	US-10-272-898-44	Sequence 44, Appl
44	765.5	31.9	472	15	US-10-354-774-44	Sequence 44, Appl
45	765.5	31.9	472	15	US-10-271-012-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-68
; Sequence 68, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-272-898-68

  Query Match             100.0%;  Score 2397;  DB 12;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-180;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 2
US-10-354-774-68
; Sequence 68, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 68:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-354-774-68

  Query Match             100.0%;  Score 2397;  DB 15;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-180;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 3
US-10-271-012-68
; Sequence 68, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 68:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-271-012-68

  Query Match             100.0%;  Score 2397;  DB 15;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-180;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 4
US-10-729-122-68
; Sequence 68, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 68:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-729-122-68

  Query Match             100.0%;  Score 2397;  DB 16;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-180;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 5
US-10-729-039-68
```

```
; Sequence 68, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 68:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 451 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-729-039-68

  Query Match             100.0%;  Score 2397;  DB 16;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-180;
  Matches  451;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGY 60

Qy         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAEVRVGDNVQLNTIYTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTN 120

Qy        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFV 180

Qy        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSK 240

Qy        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPD 300

Qy        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQ 360

Qy        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420

Qy        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |||||||||||||||||||||||||||||||
Db        421 YTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451


RESULT 6
US-10-272-898-66
; Sequence 66, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 66:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1276 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-272-898-66

  Query Match             93.2%;  Score 2235;  DB 12;  Length 1276;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-166;
  Matches  425;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWVE 451
                 ||||||
Db       1271 DPGWVE 1276
```

RESULT 7
US-10-354-774-66
; Sequence 66, Application US/10354774
; Publication No. US20030215468A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/354,774
; FILING DATE: 30-Jan-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-354-774-66

Query Match 93.2%; Score 2235; DB 15; Length 1276;
Best Local Similarity 99.8%; Pred. No. 1.1e-166;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970
```

```
Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWVE 451
                 ||||||
Db       1271 DPGWVE 1276
```

RESULT 8
US-10-271-012-66
; Sequence 66, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-271-012-66

Query Match 93.2%; Score 2235; DB 15; Length 1276;
Best Local Similarity 99.8%; Pred. No. 1.1e-166;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy         206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy         266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy         326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy         386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy         446 DPGWVE 451
               ||||||
Db        1271 DPGWVE 1276
```

RESULT 9
US-10-452-024-164
; Sequence 164, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-164

Query Match 93.2%; Score 2235; DB 15; Length 1276;
Best Local Similarity 99.8%; Pred. No. 1.1e-166;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy         206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy         266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy         326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy         386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy         446 DPGWVE 451
               ||||||
Db        1271 DPGWVE 1276
```

RESULT 10
US-10-729-122-66
; Sequence 66, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-729-122-66

Query Match 93.2%; Score 2235; DB 16; Length 1276;
Best Local Similarity 99.8%; Pred. No. 1.1e-166;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910
```

```
Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy         206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy         266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy         326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy         386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy         446 DPGWVE 451
               ||||||
Db        1271 DPGWVE 1276
```

```
RESULT 11
US-10-729-039-66
; Sequence 66, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 66:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1276 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-10-729-039-66
```

```
  Query Match             93.2%;  Score 2235;  DB 16;  Length 1276;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-166;
  Matches  425;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy         206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy         266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy         326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy         386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy         446 DPGWVE 451
               ||||||
Db        1271 DPGWVE 1276
```

```
RESULT 12
US-10-452-024-4
; Sequence 4, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
;   LENGTH: 1276
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-4
```

```
  Query Match             93.1%;  Score 2231;  DB 15;  Length 1276;
  Best Local Similarity   99.5%;  Pred. No. 2.2e-166;
  Matches  424;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
```

```
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db.      1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWVE 451
              ||||||
Db       1271 DPGWVE 1276
```

RESULT 13
US-10-452-024-168
; Sequence 168, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 168
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-168

Query Match 93.1%; Score 2231; DB 15; Length 1276;
Best Local Similarity 99.5%; Pred. No. 2.2e-166;
Matches 424; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWVE 451
              ||||||
Db       1271 DPGWVE 1276
```

RESULT 14
US-10-205-516-8
; Sequence 8, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-8

Query Match 93.1%; Score 2231; DB 15; Length 1276;
Best Local Similarity 99.5%; Pred. No. 2.2e-166;
Matches 424; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWVE 451
              ||||||
Db       1271 DPGWVE 1276
```

RESULT 15
US-10-205-516-22
; Sequence 22, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum

```
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
;   LENGTH: 1286
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-205-516-22

  Query Match             93.1%;  Score 2231;  DB 15;  Length 1286;
  Best Local Similarity   99.5%;  Pred. No. 2.2e-166;
  Matches  424;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 914

Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         915 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 974

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         975 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1034

Qy         206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1035 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1094

Qy         266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
               |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db        1095 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1154

Qy         326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1155 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1214

Qy         386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1215 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1274

Qy         446 DPGWVE 451
               ||||||
Db        1275 DPGWVE 1280


Search completed: September 10, 2004, 04:58:17
Job time : 89.2996 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 7.96759 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 2397
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2231	93.1	1276	2	S11455	botulinum neurotox
2	801.5	33.4	1291	2	S46431	botulinum neurotox
3	801.5	33.4	1291	2	A49777	botulinum neurotox
4	782.5	32.6	1285	2	S70582	botulinum neurotox
5	620.5	25.9	1291	1	A48940	bontoxilysin (EC 3
6	615.5	25.7	1291	2	I40631	non-proteolytic bo
7	610	25.4	1297	2	S39791	neurotoxin - Clost
8	591	24.7	1296	2	I40645	botulinum neurotox
9	588	24.5	1296	1	BTCLAB	bontoxilysin (EC 3
10	575.5	24.0	1268	2	S33411	botulinum neurotox
11	540	22.5	1252	2	S21178	botulinum neurotox
12	536.5	22.4	1274	2	I40813	neurotoxin type F
13	533.5	22.3	1251	2	JH0256	botulinum neurotox
14	401.5	16.8	1315	1	BTCLTN	tentoxilysin (EC 3
15	252	10.5	366	2	S48110	neurotoxin type F
16	243.5	10.2	369	2	S48109	neurotoxin type F
17	231.5	9.7	367	2	S48106	neurotoxin type E
18	196.5	8.2	1193	2	S68218	botulinum neurotox
19	195.5	8.2	1165	2	I40644	botulinum neurotox
20	188.5	7.9	1162	2	I40817	botulinum toxin no
21	186	7.8	1162	2	A47708	progenitor toxin n
22	182.5	7.6	1193	2	JC4901	nontoxic-nonhemagg
23	173	7.2	1196	2	JQ1467	toxin, nontoxic co
24	173	7.2	1196	2	S46430	botulinum neurotox
25	161.5	6.7	1413	2	B82877	conserved hypothet
26	161	6.7	883	2	T18484	hypothetical prote
27	156.5	6.5	2485	1	H71621	serine/threonine-s
28	155.5	6.5	1193	2	G71605	hypothetical prote
29	154.5	6.4	960	2	S72284	DNA-directed RNA p
30	154.5	6.4	1024	1	RNZQBF	DNA-directed RNA p
31	152.5	6.4	2136	2	A05037	hypothetical prote
32	150	6.3	588	2	A71661	hypothetical prote
33	149	6.2	2539	2	B71619	hypothetical prote
34	146.5	6.1	1306	2	T28313	ORF MSV152 probabl
35	146.5	6.1	2206	2	G71611	hypothetical prote
36	146.5	6.1	4550	2	T18440	hypothetical prote
37	146	6.1	608	2	T28301	ORF MSV140 hypothe
38	145	6.0	2364	2	I40884	cytotoxin L - Clos
39	144.5	6.0	1284	2	G82897	conserved hypothet
40	144	6.0	476	2	A28439	endonuclease SceI
41	143.5	6.0	2010	2	B71616	phosphatase (acid
42	142	5.9	888	2	A38539	p101 protein precu
43	141	5.9	480	2	C97121	p-loop ATPase doma
44	141	5.9	2339	2	A45597	DNA-directed RNA p
45	141	5.9	2500	2	G71609	hypothetical prote

ALIGNMENTS

RESULT 1
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegl
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 93.1%; Score 2231; DB 2; Length 1276;
Best Local Similarity 99.5%; Pred. No. 1.5e-119;
Matches 424; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy          86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy         146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy         206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy         266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
               ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy         326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy         386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy         446 DPGWVE 451
               ||||||
```

```
Db       1271 DPGWVE 1276


RESULT 2
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

  Query Match             33.4%;  Score 801.5;  DB 2;  Length 1291;
  Best Local Similarity   41.7%;  Pred. No. 5.3e-38;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :|||||||||||:||||||||||:|| ||||||||||     :|||| |:  |||| ||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIPPFDFKLGSS 914

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 973

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:       :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        974 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| | |::  |         :|  :|||||:|||  :||||::
Db       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
              ||:|||||| |::: |||::| :|::||:   |  :|      : :     |  | ||
Db       1093 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 1152

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
              :      :|:  || :    |  |:: | :    :|:||       |   :||: |:    :
Db       1153 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 1209

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
              || |: : :   | |||| |:|  ||   :  |   :||      ::: |     |
Db       1210 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 1268

Qy        427 T--NYETKLLSTSSFWKFI 443
              || : | |||: | |:
Db       1269 KQGNYASLLESTSTHWGFV 1287


RESULT 3
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and who
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             33.4%;  Score 801.5;  DB 2;  Length 1291;
  Best Local Similarity   41.7%;  Pred. No. 5.3e-38;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :|||||||||||:||||||||||:|| ||||||||||     :|||| |:  |||| ||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 973

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:       :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        974 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| | |::  |         :|  :|||||:|||  :||||::
Db       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
              ||:|||||| |::: |||::| :|::||:   |  :|      : :     |  | ||
Db       1093 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 1152

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
              :      :|:  || :    |  |:: | :    :|:||       |   :||: |:    :
Db       1153 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 1209

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
              || |: : :   | |||| |:|  ||   :  |   :||      ::: |     |
Db       1210 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 1268

Qy        427 T--NYETKLLSTSSFWKFI 443
              || : | |||: | |:
Db       1269 KQGNYASLLESTSTHWGFV 1287


RESULT 4
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
```

A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membr
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

Query Match 32.6%; Score 782.5; DB 2; Length 1285;
Best Local Similarity 40.4%; Pred. No. 6.4e-37;
Matches 188; Conservative 73; Mismatches 123; Indels 81; Gaps 14;

```
Qy          26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
               :||||:||||||||||||||||||||| |:|||||||||||  ||||| |:  |||| ||
Db         851 SLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSS 910

Qy          86 GD---KIIVNLNNNILYSAIYENSSVSPWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
               ||   ||||  | ||:|:|:||: |:||||:|:| ::|    ||||:|:: |||| : |
Db         911 GDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 969

Qy         143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
               :  : : |:      : : | |  | :  || |||||||| |:|| | :|||:|  :
Db         970 SNFLVFTLKQNENSEQDINFSYDISKNAAGY-NKWFFVTITTNMMGNMMIYINGKLIDTI 1028

Qy         203 KIEDLDEVKLDKTIVF--------GIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
               |:::|  :   ||| |        |:  :  :|  :||||| ||:||| ::||||::
Db        1029 KVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSL 1088

Qy         255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
                ||:|||||| |::| |||:|| ||::||::: : | :|     : :    |  | ||
Db        1089 QYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFNEGYKIIIKR 1148

Qy         315 VSDKNPYSRILNGDNII------------LHMLYNS--------------------RKY- 341
               :      :|:  |:|::             | |   |                      |||
Db        1149 IRGNTNDTRV-RGENVLYFNTTIDNKQYSLGMYKPSRNLGTDLVPLGALDQPMDEIRKYG 1207

Qy         342 -MIIRDTDTI--YATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSS 398
                 ||:  :|   ||:|             | | ||     :|| || :              |
Db        1208 SFIIQPCNTFDYYASQ-----------LFLSSNATTNRLGILSIGSY------------S 1244

Qy         399 FRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFI 443
               |:    | | |  | |: :     : : :| : | |||: | |:
Db        1245 FK-----LGDDY--W-FNHEYLIPVIKIEHYASLLESTSTHWVFV 1281
```

RESULT 5
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this public
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific i
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisi
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars,
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus t
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 25.9%; Score 620.5; DB 1; Length 1291;

```
  Best Local Similarity   34.0%;  Pred. No. 1.1e-27;
  Matches  160;  Conservative   76;  Mismatches  163;  Indels   71;  Gaps   13;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
Db        848 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA 905

Qy         87 D-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLES 1024

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVI 260
Db       1025 NTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYL 1084

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRSK------LYTGNP 309
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 1144

Qy        310 ITIKSVSDKNPYSR--ILNGDNIILHMLYNSRKYMI----------------IRDTDTI 350
Db       1145 FIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEF 1204

Qy        351 YAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRE 401
Db       1205 YNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV-----------FEE 1252

Qy        402 NTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
Db       1253 --------YKDYFCISKWYLKEVKRKPYNLKL---GCNWQFIPKDEGWTE 1291


RESULT 6
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match          25.7%;  Score 615.5;  DB 2;  Length 1291;
  Best Local Similarity   33.0%;  Pred. No. 2.2e-27;
  Matches  158;  Conservative   88;  Mismatches  152;  Indels   81;  Gaps   16;

Qy         23 ASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL 82
Db        844 SNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKL 901

Qy         83 SSSGD-KIIVNLNNNILYSAIYENSSVSFWIKISK----DLTN-SHNEYTIINSIEQNSG 136
Db        902 TSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 961

Qy        137 WKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYING 196
Db        962 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 1020

Qy        197 ELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQIL 256
Db       1021 TLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSY 1080

Qy        257 RNVIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRS------KLY 305
Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLY 1140

Qy        306 TGNPITIKSVSDKNPYSRILNGD----------NIILH----MLYNSRKYM--------- 342
Db       1141 IGEKFIIRRESN----SQSINDDIVRKEDYIHLDLVLHHEEWRVY-AYKYFKEQEEKLFL 1195

Qy        343 -IIRDTDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYC 392
Db       1196 SIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKK--- 1252

Qy        393 SQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
Db       1253 ----------------YKDYFCISKWYLKEVKRKPYKSNL---GCNWQFIPKDEGWTE 1291


RESULT 7
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match          25.4%;  Score 610;  DB 2;  Length 1297;
  Best Local Similarity   31.1%;  Pred. No. 4.5e-27;
  Matches  146;  Conservative   79;  Mismatches  176;  Indels   68;  Gaps   11;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS- 85
Db        853 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSE 912

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
Db        913 NSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVS 972

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
```

```
Db        973 IKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKK 1032

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260

Db       1033 SEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTL 1092

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----------APESNV-LVLVRYPDRSKLYTGN 308

Db       1093 KDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINY---QNLYLGL 1149

Qy        309 PITIKSVSDKNPYSRILNGDNII-------------------LHMLYNSRKYMI------ 343

Db       1150 RFIIKKASN----SRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAP 1205

Qy        344 IRDTDTIY-ATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFREN 402

Db       1206 INDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--- 1262

Qy        403 TMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451

Db       1263 ------------FCISQWYLRRISENINKLRLGCN--WQFIPVDEGWTE 1297


RESULT 8
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             24.7%;  Score 591;  DB 2;  Length 1296;
  Best Local Similarity   30.1%;  Pred. No. 5.5e-26;
  Matches  144;  Conservative   82;  Mismatches  152;  Indels  100;  Gaps   12;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL-SSSG 86

Db        862 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 921

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLT--NSHNEYTIINSIEQNSGWKLCIRNG 144

Db        922 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKINLNNEYTIINCIENNSGWKVSLNYG 981

Qy        145 NIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKI 204

Db        982 EIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQKPI 1041

Qy        205 EDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYW 264

Db       1042 SNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILKDFW 1101

Qy        265 GNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGNPI 310

Db       1102 GNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLN-STLYEGTKF 1160

Qy        311 TIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGG------------ 356

Db       1161 IIKKYASGNEDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAGVEKILSALEIPDVG 1215

Qy        357 ------------------ECSQNCVYALKLQSNLGN----YGIGIF-SIKNIVSKNKYCS 393

Db       1216 NLSQVVVMKSKDDQGIRNKCKMN------LQDNNGNDIGFIGFHLYDNIAKLVASNWYNR 1269

Qy        394 QIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451

Db       1270 QVGKASR----------------------------------TFGCSWEFIPVDDGWGE 1293


RESULT 9
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxir
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
```

J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 24.5%; Score 588; DB 1; Length 1296;
Best Local Similarity 31.6%; Pred. No. 8.1e-26;
Matches 142; Conservative 92; Mismatches 173; Indels 42; Gaps 14;

```
Qy          28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FKLSS 84
               |      ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    | | |
Db         862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy          85 SGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKLCIR 142
               |  || | | | |:|:::||| | ||||:| |  :  : :|||||| :| |||||: :
Db         922 S--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLN 979

Qy         143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
               | | | |||     : ::| ||: :: : | |:| ||||||| :  |:|||| |   :
Db         980 YGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQK 1039

Qy         203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
               | :|  :    |:| :|   | :: :||: ||:| |||: ::|  :|: |    ::||
Db        1040 PISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKD 1099

Qy         263 YWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGN 308
               :||: |::|  ||::|    : |:|          |:  |  :|:    | : | || |
Db        1100 FWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGT 1158

Qy         309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
               ||  :  |  : : | | : :::: :::| :  :     |:| |   :  : ||::
Db        1159 KFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSALEI 1211

Qy         369 QSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVT 427
                 ::||   : :   ||       |          | :     ::     | |    || :
Db        1212 -PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLVA-S 1264

Qy         428 NYETKLLSTSS-----FWKFISRDPGWVE 451
               |:  : :  ||      |:||  | || |
Db        1265 NWYNRQIERSSRTLGCSWEFIPVDDGWGE 1293
```

RESULT 10
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 24.0%; Score 575.5; DB 2; Length 1268;
Best Local Similarity 28.5%; Pred. No. 4.1e-25;
Matches 133; Conservative 95; Mismatches 155; Indels 83; Gaps 12;

```
Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
               :|  |:  ::  | || ||::: : |  :|:||| : : :  :: : :   | | : ||
Db         845 ILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYSSR 904

Qy          87 -DKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSI-EQNSGWKLCIR 142
                :: :  || |:|:: |:| |||||:| |  :| | :|||||||  :   |||||: :
Db         905 LSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNNLKNLNNEYTIINCMRNNNSGWKISLN 964

Qy         143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
               || | |||     : |:|:|:: :  : | ||| ||||||| :|: ||||||  |   :
Db         965 YNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGNLTDQK 1024

Qy         203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
               | :|  : :|  |:| |    :: : :  || | ||: ||   :|  :|  :    ::||
Db        1025 SILNLGNIHVDDNILFKI-VGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDSTILKD 1083

Qy         263 YWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYP---------DRSKLYTGNPITIK 313
               :||| | :: :|::|     :  :    |::| : |                ::||||  : |:
Db        1084 FWGNYLLYNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNIFSNARLYTGVEVIIR 1143

Qy         314 SVSD---KNPYSRILNGDNIILHMLYNSRKYMIIRDTDT---------------------- 349
               |       |  : :   | : ::::  : :| :  |  |
Db        1144 KVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRISNSNYNSNQ 1203

Qy         350 --IYATQGGECSQNCVYALKLQSNLGNYGIGI--FSIKNIVSKNKYCSQIFSSFRENTML 405
                 | : |  |: |       ::| ||  ||:  | : |:|: : |    : : | ||
Db        1204 MIIMDSIGDNCTMN------FKTNNGN-DIGLLGFHLNNLVASSWY----YKNIRNNTR- 1251

Qy         406 LADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
                                          :   || |||:: || |
Db        1252 ----------------------------NNGCFWSFISKEHGWQE 1268
```

RESULT 11
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin.
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960

A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

Query Match 22.5%; Score 540; DB 2; Length 1252;
Best Local Similarity 31.3%; Pred. No. 4.2e-23;
Matches 142; Conservative 84; Mismatches 163; Indels 64; Gaps 18;

```
Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDF-----K 81
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKI---SKDLTNSHNEYTIINSI-EQNSGW 137
Db        895 LS----EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGW 950

Qy        138 KLCIRNGNIEWILQD---VNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYI 194
Db        951 KVSLNHNEIIWTLQDNAGINQK---LAFNYGNANGISDYINKWIFVTITNDRLGDSKLYI 1007

Qy        195 NGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
Db       1008 NGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE 1066

Qy        255 ILRNVIKDYWGNPLKFDTEYYIIN----DNYIDRYIAP-------ESNVLVLVRYPDRSK 303
Db       1067 PNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILL------ANR 1120

Qy        304 LYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYM-IIRDTDTIYATQGGECSQNC 362
Db       1121 LYSGIKVKIQRVNNSS------TNDNLVRK---NDQVYINFVASKTHLFPLYADTATTNK 1171

Qy        363 VYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYT 422
Db       1172 EKTIKISSSGNRFNQVV--VMNSVGNN--CTMNFKNNNGNNIGL--------LGFK-ADT 1218

Qy        423 PVAVTNYETKLL----STSSFWKFISRDPGWVE 451
Db       1219 VVASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1251
```

RESULT 12
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific i
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 22.4%; Score 536.5; DB 2; Length 1274;
Best Local Similarity 30.2%; Pred. No. 6.9e-23;
Matches 139; Conservative 79; Mismatches 156; Indels 87; Gaps 15;

```
Qy         33 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG-DKIIV 91
Db        859 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 918

Qy         92 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-EQNSGWKL---CIRNGN 145
Db        919 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 978

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
Db        979 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 1038

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
Db       1039 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 1097

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVL------------VLVRYPDRSKLYTG------ 307
Db       1098 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 1153

Qy        308 --NPITIKSVSD---KNPYSRILNGDNIILHMLY---NSRKYMIIRDTD--------TIY 351
Db       1154 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 1213

Qy        352 ATQGGECSQNCVYALKLQSNLG-NYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIY 410
Db       1214 DSIGNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS------ 1257

Qy        411 KPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
Db       1258 ------------------------SNGCFWSSISKENGWKE 1274
```

RESULT 13
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum typ
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 22.3%; Score 533.5; DB 2; Length 1251;
Best Local Similarity 30.8%; Pred. No. 9.9e-23;
Matches 139; Conservative 86; Mismatches 165; Indels 61; Gaps 14;

```
Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDF-----K 81
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKI---SKDLTNSHNEYTIINSI-EQNSGW 137
Db        895 LS----EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGW 950

Qy        138 KLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGE 197
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        198 LKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILR 257
Db       1011 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        258 NVIKDYWGNPLKFDTEYYIIN----DNYIDRYIAP-------ESNVLVLVRYPDRSKLYT 306
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILL------ANRLYS 1123

Qy        307 GNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYAL 366
Db       1124 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADT--------ATTNKEKTI 1175

Qy        367 KLQS--NLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPV 424
Db       1176 KISSSGNRFNQVVVMNSVGN-------CTMNFKNNNGNNIGL--------LGFK-ADTVV 1219

Qy        425 AVTNYETKLL----STSSFWKFISRDPGWVE 451
Db       1220 ASTWYYTHMRDNTNSNGFFWNFISEEHGWQE 1250
```

RESULT 14
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal h
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of syn
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 16.8%; Score 401.5; DB 1; Length 1315;
Best Local Similarity 24.8%; Pred. No. 3.6e-15;
Matches 118; Conservative 92; Mismatches 158; Indels 107; Gaps 16;

```
Qy         39 INDSKILSLQNKKNALVDTSGYNAEVRVGDNVQL------NTIYTNDFKLSSSGDKIIVN 92
Db        881 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHL----VNNESSEVIVH 936

Qy         93 LNNNILYSAIYENSSVSFWIKISKDLTNSH------NEYTIINSIEQN-----SGWKLCI 141
Db        937 KAMDIEYNDMFNNFTVSFWLRVPK-VSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSL 995

Qy        142 RNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQS 201
Db        996 KGNNLIWTLKDSAGEVRQITFRDLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGS 1055

Qy        202 QKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIK 261
Db       1056 AEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLR 1115

Qy        262 DYWGNPLKFDTEYYII--NDNYIDRYIAPESNVLVLVRYPDRS---------KLYTGNPI 310
```

```
Db       1116 DFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKF 1175

Qy        311 TIKSVSDKNPY-SRILNGDNIILHMLYNSRKYMI-------------------------- 343
              ||  :  |   | : :|| | |:: ||: ::::
Db       1176 IIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIP 1235

Qy        344 ---------IRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQ 394
                     :||  | |: |           |||  :  |  :|:    |         |
Db       1236 LYKKMEAVKLRDLKT-YSVQ----------LKLYDD-KNASLGLVGTHN--------GQ 1274

Qy        395 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 449
              | :    : :: ::    | |:           : : |:|     | |:  | ||
Db       1275 IGNDPNRDILIASN----WYFN----------HLKDKILGCD--WYFVPTDEGW 1312


RESULT 15
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             10.5%;  Score 252;  DB 2;  Length 366;
  Best Local Similarity   34.8%;  Pred. No. 2.2e-07;
  Matches   49;  Conservative   35;  Mismatches   53;  Indels    4;  Gaps    3;

Qy         33 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL-SSSGDKIIV 91
              |: :  | |: || :: : |  :| ||| : : :  :| : :   | | : ||   :: :
Db        226 NKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNI 285

Qy         92 NLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSI-EQNSGWKLCIRNGNIEW 148
                ||:|:|:  |:| |:|||::| |  :  | :||||||: |   |||||: :    | |
Db        286 AQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNKIIW 345

Qy        149 ILQDVNRKYKSLIFDYSESLS 169
              |||     : |:|:|:: :|
Db        346 TLQDTAGNNQKLVFNYTQMIS 366


Search completed: September 10, 2004, 03:44:15
Job time : 8.96759 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 5.5028 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 2397
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2231	93.1	1276	1	BXD_CLOBO	P19321 clostridium
2	801.5	33.4	1290	1	BXC1_CLOBO	P18640 clostridium
3	620.5	25.9	1290	1	BXB_CLOBO	P10844 clostridium
4	610	25.4	1296	1	BXG_CLOBO	Q60393 clostridium
5	591	24.7	1295	1	BXA2_CLOBO	Q45894 clostridium
6	588	24.5	1295	1	BXA1_CLOBO	P10845 clostridium
7	536.5	22.4	1274	1	BXF_CLOBO	P30996 clostridium
8	533.5	22.3	1250	1	BXE_CLOBU	P30995 clostridium
9	529.5	22.1	1250	1	BXE_CLOBO	Q00496 clostridium
10	401.5	16.8	1314	1	TETX_CLOTE	P04958 clostridium
11	188.5	7.9	1162	1	BXEN_CLOBU	Q06366 clostridium
12	186	7.8	1162	1	BXEN_CLOBO	P46082 clostridium
13	173	7.2	1196	1	BXCN_CLOBO	P46081 clostridium
14	155	6.5	1111	1	EX5C_BUCBP	Q89ab4 buchnera ap
15	154.5	6.4	1024	1	RPOB_PLAFA	P21421 plasmodium
16	152.5	6.4	2136	1	YCF2_MARPO	P09975 marchantia
17	145	6.0	802	1	CSD2_ECOLI	P53513 escherichia
18	144	6.0	476	1	RF3_SACBA	P05512 saccharomyc
19	141	5.9	2339	1	RPC1_PLAFA	P27625 plasmodium
20	140	5.8	760	1	ETF2_MSEPV	Q9yw29 melanoplus
21	140	5.8	819	1	CSD1_ECOLI	P53512 escherichia
22	137.5	5.7	1061	1	EX5C_BUCAP	Q8k9b0 buchnera ap
23	132.5	5.5	626	1	HTPG_BUCBP	Q89a93 buchnera ap
24	131	5.5	1956	1	ATX1_PLAFA	Q04956 plasmodium
25	130.5	5.4	799	1	Y231_BUCAP	Q8k9s5 buchnera ap
26	130.5	5.4	2894	1	YD96_METJA	Q58791 methanococc
27	130	5.4	351	1	AROB_CAMJE	Q9pnt2 campylobact
28	128	5.3	386	1	RMAR_HANWI	P48849 hansenula w
29	127.5	5.3	519	1	YC82_ASTLO	P58145 astasia lon
30	127	5.3	559	1	ENS2_YEAST	P12294 saccharomyc
31	127	5.3	789	1	AI2M_YEAST	P03876 saccharomyc
32	126.5	5.3	537	1	YA60_METJA	Q58460 methanococc
33	126.5	5.3	964	1	DPOL_CBEPV	P30319 choristoneu
34	126	5.3	501	1	AMPA_WIGBR	Q8d295 wiggleswort
35	124	5.2	1099	1	PLC1_CANAL	O13433 candida alb
36	122.5	5.1	552	1	YC92_METJA	Q58688 methanococc
37	122.5	5.1	638	1	HTPG_WIGBR	Q8d228 wiggleswort
38	121.5	5.1	1361	1	YME9_YEAST	Q04693 saccharomyc
39	121	5.0	667	1	Y366_MYCGE	P47606 mycoplasma
40	121	5.0	878	1	SECA_ANTSP	Q06461 antithamnio
41	120.5	5.0	679	1	PAN3_YEAST	P36102 saccharomyc
42	120.5	5.0	1120	1	MFD_RICPR	O05955 rickettsia
43	120.5	5.0	1197	1	DPOM_PODAN	Q01529 podospora a
44	120	5.0	735	1	CIGB_DICDI	Q94481 dictyosteli
45	120	5.0	854	1	CC24_YEAST	P11433 saccharomyc

ALIGNMENTS

RESULT 1
BXD_CLOBO
ID BXD_CLOBO STANDARD; PRT; 1276 AA.
AC P19321;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE (Bontoxilysin D).
GN BOTD.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BVD/-3;
RX MEDLINE=91016853; PubMed=2216736;
RA Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT "Nucleotide sequence of the gene encoding Clostridium botulinum
RT neurotoxin type D.";
RL Nucleic Acids Res. 18:5556-5556(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CB16;
RX MEDLINE=93042276; PubMed=1420572;
RA Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT "The complete amino acid sequence of the Clostridium botulinum type D
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT phage d-16 phi genome.";
RL J. Vet. Med. Sci. 54:905-913(1992).
RN [3]
RP PARTIAL SEQUENCE.
RC STRAIN=D-SA, and D-1873;
RX MEDLINE=89339741; PubMed=2668193;
RA Moriishi K., Syuto B., Kubo S., Oguma K.;
RT "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT strains.";
RL Infect. Immun. 57:2886-2891(1989).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinal neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC SYNAPTOBREVINS-1 AND -2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

```
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match           93.1%;  Score 2231;  DB 1;  Length 1276;
  Best Local Similarity   99.5%;  Pred. No. 9.5e-121;
  Matches  424;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWVE 451
              ||||||
Db       1271 DPGWVE 1276


RESULT 2
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
```

```
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             33.4%;  Score 801.5;  DB 1;  Length 1290;
  Best Local Similarity   41.7%;  Pred. No. 1.1e-38;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||:||||||||||:|| |||||||||||    :|||| |:  |||| ||
Db        854 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 913

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        914 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 972

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        973 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 1031

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| | |::  |        :|  :|||| ||:|||  :||||::
Db       1032 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1091

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
                ||:|||||| |::: |||::| :|::||:   |  :|     : :    |  | ||
Db       1092 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 1151

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
              :      :|:  || :   |  |:: | :    :|:||      |   :||: |:    :

Db       1152 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 1208

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
                 || |: : :   | |||| |:|  ||   :  |   :||      ::: |    |
Db       1209 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 1267

Qy        427 T--NYETKLLSTSSFWKFI 443
               || : | |||: | |:
Db       1268 KQGNYASLLESTSTHWGFV 1286


RESULT 3
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
```

```
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             25.9%;  Score 620.5;  DB 1;  Length 1290;
  Best Local Similarity   34.0%;  Pred. No. 2.7e-28;
  Matches  160;  Conservative   76;  Mismatches  163;  Indels   71;  Gaps   13;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              :| :: |:| : | :: ||:|: | | |:| ||| |:| | | |:||    | |||:||
Db        847 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSA 904

Qy         87 D-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
              : || |  | ||:::::: : ||||||:| |   :     ||||||||| :: |||||:
Db        905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 964

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              ||   | | | |:| | ||: |:|:     : | |:|||||||||| :   |:||||:|:
Db        965 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLES 1023

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVI 260
              :  |:|: ||  :  |:| :| :||  | :|:: |:||: |||  :|   |: |     :
Db       1024 NTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYL 1083

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRSK------LYTGNP 309
              ||:||||| :: |||: |    : ||     :|   :|   :|   ||       || |
Db       1084 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 1143

Qy        310 ITIKSVSDKNPYSR--ILNGDNIILHMLYNSRKYMI----------------IRDTDTI 350
                |:  |:    :   :   | | |     :::: :                | |:|
Db       1144 FIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISDSDEF 1203

Qy        351 YAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRE 401
              | | |  |  :   |: :|           :|  ||  |    ||          | |
Db       1204 YNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV-----------FEE 1251

Qy        402 NTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
                  || :    |     |    |  ||      |:|| :| || |
Db       1252 --------YKDYFCISKWYLKEVKRKPYNLKL---GCNWQFIPKDEGWTE 1290


RESULT 4
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
```

```
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             25.4%;  Score 610;  DB 1;  Length 1296;
  Best Local Similarity   31.1%;  Pred. No. 1.1e-27;
  Matches  146;  Conservative   79;  Mismatches  176;  Indels   68;  Gaps   11;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS- 85
              :|  : | | ::|: : ||||  :   |:|:||| | : || :|  | |    |||::|
Db        852 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSE 911

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
                 |  : :  ::| ::::| |::||::  |   |       ||||||: |: :||||:
Db        912 NSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVS 971

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              |:   | | | ||| | ||: |:||   : : | |||| :||||: :|   :|||| ||:
Db        972 IKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKK 1031

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRNVI 260
              |:|| :|| :     | | :    |  : :||:||||| :||:  ::: :|  |   | :
Db       1032 SEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTL 1091

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----------APESNV-LVLVRYPDRSKLYTGN 308
              ||:|||||::||:||: |    : ||           || :|     : |     || |
Db       1092 KDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINY---QNLYLGL 1148

Qy        309 PITIKSVSDKNPYSRILNGDNII------------------LHMLYNSRKYMI------ 343
                 ||  |:    || :| |||:                  :::| ||::
Db       1149 RFIIKKASN----SRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAP 1204

Qy        344 IRDTDTIY-ATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFREN 402
              | |  | |    |  :  :   | ::          |:| |   |     |  : ::
Db       1205 INDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--- 1261

Qy        403 TMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
                        |     |       |     |  :  |:||  | || |
Db       1262 ------------FCISQWYLRRISENINKLRLGCN--WQFIPVDEGWTE 1296


RESULT 5
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
```

```
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             24.7%;  Score 591;  DB 1;  Length 1295;
  Best Local Similarity   30.1%;  Pred. No. 1.3e-26;
  Matches  144;  Conservative   82;  Mismatches  152;  Indels  100;  Gaps   12;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL-SSSG 86
Db        861 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 920

Qy         87 DKIIVNLNNNILYSAIYENSSVSFWIKISKDLT--NSHNEYTIINSIEQNSGWKLCIRNG 144
Db        921 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKINLNNEYTIINCIENNSGWKVSLNYG 980

Qy        145 NIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKI 204
Db        981 EIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQKPI 1040

Qy        205 EDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYW 264
Db       1041 SNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILKDFW 1100

Qy        265 GNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGNPI 310
Db       1101 GNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLN-STLYEGTKF 1159

Qy        311 TIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGG-------------- 356
Db       1160 IIKKYASGNEDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAGVEKILSALEIPDVG 1214

Qy        357 ------------------ECSQNCVYALKLQSNLGN----YGIGIF-SIKNIVSKNKYCS 393
Db       1215 NLSQVVVMKSKDDQGIRNKCKMN------LQDNNGNDIGFIGFHLYDNIAKLVASNWYNR 1268

Qy        394 QIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
Db       1269 QVGKASR---------------------------------TFGCSWEFIPVDDGWGE 1292


RESULT 6
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
```

```
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             24.5%;  Score 588;  DB 1;  Length 1295;
  Best Local Similarity   31.6%;  Pred. No. 2e-26;
  Matches  142;  Conservative   92;  Mismatches  173;  Indels   42;  Gaps   14;

Qy         28 LKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTND---FKLSS 84
              |     ||  :| :: ||:|: : | |:| | | ::: :|  |  : |  |    | | |
Db        861 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 920

Qy         85 SGDKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSIEQNSGWKLCIR 142
              |  || | | |:|:::||| | ||||:| |  :  : :|||||||  :| |||||: :
Db        921 S--KIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLN 978

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              | | |||     : ::| ||: :: : | |:| |||||||  :   |:|||| |   :
Db        979 YGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQK 1038

Qy        203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
              | :|  :     |:| :|   | :: :||: ||:| |||: ::|  :|: |    ::||
Db       1039 PISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKD 1098

Qy        263 YWGNPLKFDTEYYIIN----DNYID---------RYI-APESNVLVLVRYPDRSKLYTGN 308
              :||: |::|  ||::|    : |:|          |:  |  :|:    | : | || |
Db       1099 FWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGT 1157

Qy        309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
              ||  :  |  : : | | : ::::  :::| :  :     |:| |   :  : ||::
Db       1158 KFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATN-----ASQAG--VEKILSALEI 1210

Qy        369 QSNLGNYG-IGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVT 427
              ::||   : :  ||       |       | :   ::     | |    || :
Db       1211 -PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQ-----FNNIAKLVA-S 1263

Qy        428 NYETKLLSTSS-----FWKFISRDPGWVE 451
              |:  : :  ||      |:||  | || |
Db       1264 NWYNRQIERSSRTLGCSWEFIPVDDGWGE 1292


RESULT 7
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
```

```
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             22.4%;  Score 536.5;  DB 1;  Length 1274;
  Best Local Similarity   30.2%;  Pred. No. 1.8e-23;
  Matches  139;  Conservative   79;  Mismatches  156;  Indels   87;  Gaps   15;

Qy         33 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG-DKIIV 91
              |  :  | || || :: : |  :| ||| : : :  || : :    | | : :|   :: :
Db        859 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 918

Qy         92 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-EQNSGWKL---CIRNGN 145
               ||:|:|:: |:| |:|||::| |      | : |||||| :   |||||:     :|: .
Db        919 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 978

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              | | ||| :   ::||| | |    : | ||| |||||||| :|  ::|||| |   : |
Db        979 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 1038

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              :| :: :   |:| |    |:   : || | :|: ||   :|  :|  :   :::|:|||
Db       1039 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 1097

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVL-----------VLVRYPDRSKLYTG------ 307
              | | :: :||: |    |:||   | :|           | : |    ||| |
Db       1098 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 1153

Qy        308 --NPITIKSVSD---KNPYSRILNGDNIILHMLY---NSRKYMIIRDTD--------TIY 351
                || | :  :   ||  : |   |  : : ||    | |  ||| ::         :
Db       1154 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 1213

Qy        352 ATQGGECSQNCVYALKLQSNLG-NYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIY 410
              : |  |: |        |:| | | |:  |   |:|: : |    ::: | ||
Db       1214 DSIGNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS------ 1257

Qy        411 KPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
                                    |   ||  ||:: || |
Db       1258 ----------------------SNGCFWSSISKENGWKE 1274


RESULT 8
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
```

```
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             22.3%;  Score 533.5;  DB 1;  Length 1250;
  Best Local Similarity   30.8%;  Pred. No. 2.6e-23;
  Matches  139;  Conservative   86;  Mismatches  165;  Indels   61;  Gaps   14;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDF-----K 81
              :|   . |::|  |  | :|::: | :  ||||||:: : :  :|      | |     |
Db        834 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 893

Qy         82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKI---SKDLTNSHNEYTIINSI-EQNSGW 137
              ||    :: :: |: |:|   |:| |:|||::|     : | :|||||||| : : ||||
Db        894 LS----EVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGW 949

Qy        138 KLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGE 197
              |: : :  | | ||| :  : | |:|  :   : | ||| ||||||: :|  ||||||
Db        950 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        198 LKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILR 257
              |  : | :|  : :   |:| |  |     : : || |||| |||   :|  :|  :
Db       1010 LIDKKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1068

Qy        258 NVIKDYWGNPLKFDTEYYIIN----DNYIDRYIAP-------ESNVLVLVRYPDRSKLYT 306
              |::||:||| | :| |||::|    :|:|:|            | :|:       ::||:
Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIRSTILL------ANRLYS 1122

Qy        307 GNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYAL 366
              |  : |: |:: :    ::  :: :      |: :::    ||        : |    :
Db       1123 GIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLLPLYADT--------ATTNKEKTI 1174

Qy        367 KLQS--NLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPV 424
              |: |  |  | : : |: |      |:  | :   | : |          || | | |
Db       1175 KISSSGNRFNQVVVMNSVGN-------CTMNFKNNNGNNIGL--------LGFK-ADTVV 1218

Qy        425 AVTNYETKLL----STSSFWKFISRDPGWVE 451
              | | | | :     |   || ||| : || |
Db       1219 ASTWYYTHMRDNTNSNGFFWNFISEEHGWQE 1249


RESULT 9
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
```

```
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             22.1%;  Score 529.5;  DB 1;  Length 1250;
  Best Local Similarity   31.0%;  Pred. No. 4.4e-23;
  Matches  141;  Conservative   84;  Mismatches  161;  Indels   69;  Gaps   18;

Qy          27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDF-----K 81
               :|    |::|  |  | :|:|:: | :  ||||||:: : :  :|       | |     |
Db         834 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 893

Qy          82 LSSSGDKIIVNLNNNILYSAIYENSSVSFWIKI---SKDLTNSHNEYTIINSI-EQNSGW 137
               ||     :: :: |: |:|   |:| |:|||::|      : | :|||||| : : ||||
Db         894 LS----EVNISQNDYIIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGW 949

Qy         138 KLCIRNGNIEWILQD---VNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYI 194
               |: : :  | |  :|   :|:|   | |:|  :  : | ||| ||||||: :|  ||||
Db         950 KVSLNHNEIIWTFEDNRGINQK---LAFNYGNANGISDYINKWIFVTITNDRLGDSKLYI 1006

Qy         195 NGELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQ 254
               || |   : | :|  : :   |:| |  |    : : || |||| |||   :|  :| :
Db        1007 NGNLIDQKSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE 1065

Qy         255 ILRNVIKDYWGNPLKFDTEYYIIN----DNYIDRYIAP-------ESNVLVLVRYPDRSK 303
               |::||:||| | :| |||::|    :|:|||             | :|:       ::
Db        1066 PNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILL------ANR 1119

Qy         304 LYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYM-IIRDTDTIYATQGGECSQNC 362
               ||:|  : |: |:: :          ||::     | : |:  :     ::       : |
Db        1120 LYSGIKVKIQRVNNSS------TNDNLVRK---NDQVYINFVASKTHLFPLYADTATTNK 1170

Qy         363 VYALKLQS--NLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNA 420
                 :|: |  |  |  : : |: |        |:  | :   | : |                | |
Db        1171 EKTIKISSSGNRFNQVVVMNSVGN-------CTMNFKNNNGNNIGL--------LGFK-A 1214

Qy         421 YTPVAVTNYETKLL----STSSFWKFISRDPGWVE 451
                 | || | | | :     |   || ||| : || |
Db        1215 DTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1249


RESULT 10
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
```

```
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).

CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
```

```
FT   TURN         969    970
FT   STRAND       972    977
FT   STRAND       980    981
FT   HELIX        983    985
FT   STRAND       987    995
FT   TURN         996    997
FT   STRAND       998   1004
FT   TURN        1006   1007
FT   STRAND      1010   1016
FT   STRAND      1020   1020
FT   TURN        1021   1022
FT   STRAND      1031   1037
FT   TURN        1039   1040
FT   STRAND      1042   1047
FT   TURN        1048   1049
FT   STRAND      1050   1056
FT   TURN        1058   1059
FT   STRAND      1068   1074
FT   TURN        1079   1080
FT   STRAND      1082   1091
FT   HELIX       1097   1105
FT   TURN        1106   1107
FT   STRAND      1112   1112
FT   STRAND      1114   1114
FT   TURN        1116   1117
FT   STRAND      1120   1120
FT   STRAND      1122   1122
FT   TURN        1123   1124
FT   STRAND      1127   1131
FT   HELIX       1132   1134
FT   TURN        1135   1136
FT   STRAND      1137   1141
FT   TURN        1144   1145
FT   STRAND      1148   1152
FT   STRAND      1155   1158
FT   TURN        1159   1162
FT   STRAND      1163   1166
FT   STRAND      1173   1178
FT   TURN        1184   1185
FT   STRAND      1188   1188
FT   STRAND      1190   1190

  Query Match             16.8%;  Score 401.5;  DB 1;  Length 1314;
  Best Local Similarity   24.8%;  Pred. No. 1e-15;
  Matches  118;  Conservative   92;  Mismatches  158;  Indels  107;  Gaps   16;

Qy         39 INDSKILSLQNKKNALVDTSGYNAEVRVGDNVQL------NTIYTNDFKLSSSGDKIIVN 92
              :  | ||:|    : : | ||:|: |    : ||       |:     :::    ::||:
Db        880 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHL----VNNESSEVIVH 935

Qy         93 LNNNILYSAIYENSSVSFWIKISKDLTNSH------NEYTIINSIEQN-----SGWKLCI 141
                 :| |: :: | :||||::: | :: ||      |||:||:|::::     ||| : :
Db        936 KAMDIEYNDMFNNFTVSFWLRVPK-VSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSL 994

Qy        142 RNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQS 201
              :  |: | |:|   : : : |       :    ||| |:||||: :    |||| | |
Db        995 KGNNLIWTLKDSAGEVRQITFRDLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGS 1054

Qy        202 QKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIK 261
              :|  |  :: |  |   :|   : || : |  | || | |: ::|  :|   :    ::
Db       1055 AEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLR 1114

Qy        262 DYWGNPLKFDTEYYII--NDNYIDRYIAPESNVLVLVRYPDRS---------KLYTGNPI 310
              |:|||||::|||||:|    :  |  :   :: : |   |  :          :|| |
Db       1115 DFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKF 1174

Qy        311 TIKSVSDKNPY-SRILNGDNIILHMLYNSRKYMI-------------------------- 343
              ||  :  |    | : :|| | |:: ||: ::::
Db       1175 IIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIP 1234

Qy        344 ---------IRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQ 394
                      :||  | |: |              |||  :  |  :|:   |          |
Db       1235 LYKKMEAVKLRDLKT-YSVQ-----------LKLYDD-KNASLGLVGTHN--------GQ 1273

Qy        395 IFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 449
              | :    : :: ::    | |:             : : |:|     | |:  | ||
Db       1274 IGNDPNRDILIASN----WYFN----------HLKDKILGCD--WYFVPTDEGW 1311


RESULT 11
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match              7.9%;  Score 188.5;  DB 1;  Length 1162;
  Best Local Similarity   19.0%;  Pred. No. 0.0015;
  Matches   64;  Conservative   82;  Mismatches  145;  Indels   45;  Gaps   12;

Qy         28 LKDIINEYFNSINDSKI----LSL----QNKKNALVDTSGYNAEVRVGDNVQLNTIY-TN 78
              :|:: | |   :   :     |||     :   | : |||| |  |    :: |  :| |
Db        781 MKNLFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGKNTLVEYPKDIGL--VYGIN 838

Qy         79 DFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWK 138
              :  :  :|    :   |:   : :  | |: ||:   ::|  :  :  :| | | | ||:
Db        839 NNAIHLTGANQNIKFTNDYFENGLTNNFSIYFWL---RNLNQNTIKSKLIGSKEDNCGWE 895

Qy        139 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTIT-NNIMGYMKLYINGE 197
              :  | : : : | |   |::    |  ::|:          | :: |: | :    : ::|:
Db        896 IYFENNGLVFNIIDSNGNEKNI---YLSNISN----KSWHYIVISINRLKDQLLIFIDNI 948

Qy        198 LKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILR 257
              |  :: |:::  :      |       : :|  ::|   :: :| ::: :|    |    :
Db        949 LVANEDIKEILNIYSSDIISL-----LSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNN 1003

Qy        258 NVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGN--------- 308
              : |:::    |:::  | : |       |: ||   : : :    : :| |
Db       1004 SYIRNFDEEILQYNRTYELFN------YVFPE---IAINKIEQNNNIYLSNNNENSLNFK 1054
```

```
Qy        309 PITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMII 344
              |: :| ::       :   | :|  :|  : ||: |
Db       1055 PLKFKLLNTNPNKQYVQKWDEVIFSVLDGTEKYLDI 1090
```

RESULT 12
BXEN_CLOBO
ID   BXEN_CLOBO STANDARD; PRT; 1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC   TOXICITY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA; 136856 MW; 96468EDDDAE0F39D CRC64;

  Query Match   7.8%; Score 186; DB 1; Length 1162;
  Best Local Similarity   19.7%; Pred. No. 0.0021;
  Matches   74; Conservative   85; Mismatches 155; Indels   62; Gaps   16;

```
Qy         28 LKDIINEYFNSINDSKI----LSL----QNKKNALVDTSGYNAEVRVGDNVQLNTIY-TN 78
              :|:: | |   :   :     |||    :   | : ||||| |   |    :: |  :|  |
Db        781 MKNLFNLYTELLIKEQTSPYELSLYAFQEQDNNVIGDTSGKNTLVEYPKDIGL--VYGIN 838

Qy         79 DFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWK 138
              :  :  :|    :   |:   : :  | |: ||:   ::|  :  :  :| | | | ||:
Db        839 NNAIHLTGANQNIKFTNDYFENGLTNNFSIYFWL---RNLKQNTIKSKLIGSKEDNCGWE 895

Qy        139 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTIT-NNIMGYMKLYINGE 197
              :   |  : : : | |   |::   |  ::|:      | :: |: | :   : ::|:
Db        896 IYFENDGLVFNIIDSNGNEKNI---YLSNISN----KSWHYIVISINRLKDQLLIFIDNI 948

Qy        198 LKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILR 257
              |  :: |:::  :     |       : :|  ::|   :: :| ::: :|   |   :
Db        949 LVANEDIKEILNIYSSDIISL-----LSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNN 1003

Qy        258 NVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTG----NPITIK 313
              : |:::   |:::  | : |      |: ||   : : :   : :|      | :  |
Db       1004 SYIRNFDEEILQYNRTYELFN------YVFPE---IAINKIEQNNNIYLSINNENNLNFK 1054

Qy        314 SVSDK----NPYSR-ILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKL 368
              :  |    ||  : :   | :|  :|  : ||: |  |:                ::|
Db       1055 PLKFKLLNTNPNKQYVQKWDEVIFSVLDGTEKYLDISTTNN--------------RIQL 1099

Qy        369 QSNLGNYGIGIFSIKN 384
                |  |    || | |
Db       1100 VDNKNN--AQIFIINN 1113
```

RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO STANDARD; PRT; 1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC   TOXICITY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA; 138740 MW; 4BD5956274D7F9C3 CRC64;

  Query Match   7.2%; Score 173; DB 1; Length 1196;
  Best Local Similarity   17.9%; Pred. No. 0.012;
  Matches   78; Conservative 107; Mismatches 159; Indels   92; Gaps   24;

```
Qy         27 LLKDIINEYFNSINDSKILSLQNK---KNALVDTSGYNAEVRVGDNVQLNTIY-TNDFKL 82
              ::|:|:: |       ::|   :|    | : | || |  ::  ::::|  :|  |   |
Db        827 VMKEILSPY-------QLLLFASKGPNSNIIEDISGKNTLIQYTESIEL--VYGVNGESL 877

Qy         83 SSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYT-IINSIEQNSGWKLCI 141
                 :  :|    : :  | :: ||::     |  ::: | :|  :   | ||::
Db        878 YLKSPNETIKFSNKFFTNGLTNNFTICFWLR----FTGKNDDKTRLIGNKVNNCGWEIYF 933

Qy        142 RNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTIT-NNIMGYMKLYINGELKQ 200
              :  : : : | |   :|:   |  ::      : |::::|: : :   : ::|| :
Db        934 EDNGLVFEIIDSNGNQESV---YLSNI----INDNWYYISISVDRLKDQLLIFINDKNVA 986

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
              :  |: :  : :  | :  :::|  ::: : ::    :::|::   |   :  : |
Db        987 NVSIDQI--LSIYSTNIISL---VNKNNSIYVEELSVLDNPITSEEVIRNYFSYLDNSYI 1041

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKN- 319
              :|   : |:::  | : |      |: ||::            ||       |:| |
Db       1042 RDSSKSLLEYNKNYQLYN------YVFPETS-----------LY--------EVNDNNK 1075

Qy        320 PYSRILNGDNIILHMLYNSRKYMIIR-DTDTIYATQGGECSQNCVYALKLQSNLGNYGIG 378
```

```
Db       1076 SYLSLKNTDGINI----SSVKFKLINIDESKVYVQKWDECI-ICVLD-----------G 1118

Qy        379 IFSIKNIVSKNKYCSQIFSSFRENTMLL---ADIYKP--WRFSFKNAYTPVA-------- 425

Db       1119 TEKYLDISPENNRI-QLVSS-KDNAKKITVNTDLFRPDCITFSYNDKYFSLSLRDGDYNW 1176

Qy        426 -VTNYETKLLSTSSFW 440

Db       1177 MICNDNNKVPKGAHLW 1192


RESULT 14
EX5C_BUCBP
ID   EX5C_BUCBP     STANDARD;      PRT;  1111 AA.
AC   Q89AB4;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exodeoxyribonuclease V gamma chain (EC 3.1.11.5).
GN   RECC OR BBP403.
OS   Buchnera aphidicola (subsp. Baizongia pistaciae).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=135842;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426901; PubMed=12522265;
RA   Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA   Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA   Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT   "Reductive genome evolution in Buchnera aphidicola.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC   -!- FUNCTION: EXHIBITS A WIDE VARIETY OF CATALYTIC ACTIVITIES
CC       INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED ENDONUCLEASE,
CC       ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE ACTIVITIES
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC       ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC       phosphooligonucleotides.
CC   -!- SUBUNIT: Consist of three subunits; recB, recC and recD (By
CC       similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE014017; AAO27115.1; -.
DR   InterPro; IPR006697; RecC.
DR   Pfam; PF04257; Exonuc_V_gamma; 1.
KW   Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; DNA repair;
KW   Complete proteome.
SQ   SEQUENCE   1111 AA;  132959 MW;  A13BC72BC5A62FB1 CRC64;

  Query Match             6.5%;  Score 155;  DB 1;  Length 1111;
  Best Local Similarity   21.8%;  Pred. No. 0.12;
  Matches  113;  Conservative   87;  Mismatches  181;  Indels  138;  Gaps   23;

Qy          9 HHHHSSGHIEGRHMASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGD 68

Db        231 HFFYATASYNDKQMFSTSFIK---FKQTQEISNNQLLSLLHSKNK-------NIIYHEIE 280

Qy         69 NVQLNTIY-----TNDFK----LSSSGDKIIVNLNNNILYSAI---------------- 102

Db        281 NFQKNTKFQSYWITYEYEQMLFLSLIKDKTI-NLNFKIKQNSLLQYLQNKILNLHSIQKK 339

Qy        103 ---YENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKS 159

Db        340 NNEQENNKILI-LKTDKSL-SIHSCHTLIREIEVLHNNLLNILNDNNDILFQDVLVISKN 397

Qy        160 L---------IFDYSE-------SLSHTGYTNKWFFVTITNNI-MGYMKLYINGELKQSQ 202

Db        398 LDTYIPYINQIFNSASSKNFIPFSINYKNIKNIEFYKTILELIDLPNIELNIN------- 450

Qy        203 KIED-------LDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYE--- 252

Db        451 KITNLLNIPTLLKKVQINEQEIPILLQIINQTEIQWEND-NILSNDWLSNDITKCWEHGL 509

Qy        253 -----GQILRNVIKDYWGNPL---KFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKL 304

Db        510 QRIFLGQIINNIDYKFWENIVPYNEFSTLYY----NIFNKLISLISLLNKWKKILSKPKF 565

Qy        305 YTGNPITIKSV--------SDKNPYSR-ILNGDNIILHM-----------------LYN 337

Db        566 LSHWDITFKKLLNDFFTDHEQKQPEAKLIIKKWTEIIHSGMQEKYTKKISITLLKNELCN 625

Qy        338 SRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFS 397

Db        626 SVSY--ISQSNYLFSGKITFCNN-----------------FTLTNIPFKIIYLIGLND 664

Qy        398 SFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLST 436

Db        665 NISSNTHESFDIYNLLKL-HPRAYDPCDEINHKNLLLKT 702


RESULT 15
RPOB_PLAFA
ID   RPOB_PLAFA     STANDARD;      PRT;  1024 AA.
AC   P21421;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN   RPOB.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BW(C10);
RX   MEDLINE=95107345; PubMed=7808472;
RA   Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RA   Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RT   "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
RT   Plasmodium falciparum.";
RL   Mol. Biochem. Parasitol. 66:221-231(1994).
RN   [2]
RP   SEQUENCE OF 328-1024 FROM N.A.
RX   MEDLINE=91187055; PubMed=2011147;
RA   Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT   "A circular DNA in malaria parasites encodes an RNA polymerase like
RT   that of prokaryotes and chloroplasts.";
RL   Mol. Biochem. Parasitol. 44:115-124(1991).
CC   -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC       of DNA into RNA using the four ribonucleoside triphosphates as
CC       substrates.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC   -!- SIMILARITY: Belongs to the RNA polymerase beta chain family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75544; CAA53232.1; -.
DR   EMBL; X52177; CAA36427.1; -.
DR   EMBL; X95275; CAA64572.1; -.
```

```
DR   PIR; S72282; RNZQBF.
DR   InterPro; IPR007121; RNA_pol_B.
DR   InterPro; IPR007645; RNA_pol_Rpb2_3.
DR   InterPro; IPR007120; RNA_pol_Rpb2_6.
DR   InterPro; IPR007641; RNA_pol_Rpb2_7.
DR   Pfam; PF04565; RNA_pol_Rpb2_3; 1.
DR   Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR   Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase.
SQ   SEQUENCE   1024 AA;  122185 MW;  B49FBDED7C9BD412 CRC64;

  Query Match             6.4%;  Score 154.5;  DB 1;  Length 1024;
  Best Local Similarity   23.7%;  Pred. No. 0.12;
  Matches  107;  Conservative   76;  Mismatches  149;  Indels  119;  Gaps   27;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              ||| |:    | || :||:    | | |:    :       : : || :|    :|
Db         74 LLKIILTLKLNFININKII----KFNILI----FILPFIYNNIIILNGLYKTCIQLFKKN 125

Qy         87 DKIIV----NLNNNILYSAIYENSSVSFWIKISK-DLTNSHNEYTI--------INSIEQ 133
              :|| :    | | ||:|  || :  :    |||| ::    | :          :|:|
Db        126 NKIFIIKFKNNNKNIIYVYIYISLGLRIIFKISKLNIDCYFNNFKFNFLILLLYLNNIYI 185

Qy        134 NSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLY 193
              |    | | | ||      :|:|   ||::| : : :: | |        |||:  :||:
Db        186 NKNISLFIYN-NI------INKKI--LIYNYIKFI-YSKYNN-------INNIIS-LKLF 227

Qy        194 INGELKQSQKIEDLDEVKLDK---------TIVFGIDENIDENQMLWIRDF---NIFSKE 241
              |               :||:|          |:| |  |       :  ||   ||::|:
Db        228 I--------------IKLNKFNNIYIINLLNILFSIKLNFS-----YYSDFYINNIYNKK 267

Qy        242 LSNEDINI---------VYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNV 292
                :   |:         ::: |:| |: :: : |         : |  ||:  |    .|:
Db        268 FYSIIDNLLIKSKKYLKIFKYQLL-NINRNIYNN-----ITLLLNNKKYIN-IILENINI 320

Qy        293 LVLVRYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYA 352
              ||:| |:          : ::|: |   :|    |:|   |  : |:::  |
Db        321 NPLVQYSDQ----------VNNLSEINQKFKI----NMITTGL--NSKFILNNDL------ 359

Qy        353 TQGGECSQNCV-YALKLQSNLG-NYGIGIFSIKNIVSKNKYCSQIFSSF----RENTMLL 406
                 | :| : |   : :| |   |:  :   ||    ||   |:       | |  ||
Db        360 ---RELPRNILGYISLINTNEGLTCGLVNYLTTNIFLNLKYLFVIYYKHIFYNRYNFKLL 416

Qy        407 ADIYKP--WRFSFKNAYTPVAVTNYETKLLS 435
               :|:    :  || | |    :   :| :|:
Db        417 LNIFNKNFYNISFNNIYLKKNINFNKTTILT 447


Search completed: September 10, 2004, 03:35:50
Job time : 8.5028 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 20.6355 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-68
Perfect score: 2397
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....KLLSTSSFWKFISRDPGWVE 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2226	92.9	1275	12	Q9QTG7	Q9qtg7 clostridium
2	2131	88.9	1280	2	Q45849	Q45849 clostridium
3	2131	88.9	1280	2	Q9LBS7	Q9lbs7 clostridium
4	2125	88.7	1280	2	Q841S3	Q841s3 clostridium
5	801.5	33.4	1291	2	Q93HT3	Q93ht3 clostridium
6	782.5	32.6	1285	2	Q45967	Q45967 clostridium
7	781.5	32.6	1285	2	Q9LBR1	Q9lbr1 clostridium
8	626.5	26.1	1291	2	Q93G71	Q93g71 clostridium
9	624.5	26.1	1291	2	Q933K0	Q933k0 clostridium
10	621.5	25.9	1291	2	Q9ZAJ8	Q9zaj8 clostridium
11	617.5	25.8	1291	2	Q8GR96	Q8gr96 clostridium
12	615.5	25.7	1291	2	Q08077	Q08077 clostridium
13	592.5	24.7	441	2	Q9X708	Q9x708 clostridium
14	575.5	24.0	1268	2	Q45851	Q45851 clostridium
15	571.5	23.8	1251	2	Q9K395	Q9k395 clostridium
16	555.5	23.2	1278	2	Q57236	Q57236 clostridium
17	537	22.4	1252	2	Q8KZM3	Q8kzm3 clostridium
18	537	22.4	1255	2	Q9FAR6	Q9far6 clostridium
19	534	22.3	1280	2	Q9ZAJ5	Q9zaj5 clostridium
20	402.5	16.8	451	2	Q9LA13	Q9la13 clostridium
21	396	16.5	1310	2	Q93N27	Q93n27 clostridium
22	312	13.0	361	2	Q45848	Q45848 clostridium
23	312	13.0	361	2	Q45846	Q45846 clostridium
24	250.5	10.5	1198	2	O06018	O06018 clostridium
25	233.5	9.7	1197	2	Q45888	Q45888 clostridium
26	231.5	9.7	367	2	Q45861	Q45861 clostridium
27	231.5	9.7	367	2	Q45862	Q45862 clostridium
28	220.5	9.2	1197	2	Q9ZAJ9	Q9zaj9 clostridium
29	218.5	9.1	1197	2	P71117	P71117 clostridium
30	217.5	9.1	1197	2	O33871	O33871 clostridium
31	213.5	8.9	1198	2	O69277	O69277 clostridium
32	209.5	8.7	1198	2	Q45893	Q45893 clostridium
33	207.5	8.7	1161	2	Q45891	Q45891 clostridium
34	204.5	8.5	1161	2	O69276	O69276 clostridium
35	202.5	8.4	1193	2	Q45880	Q45880 clostridium
36	196.5	8.2	1193	2	P71107	P71107 clostridium
37	196.5	8.2	1193	2	Q45914	Q45914 clostridium
38	195.5	8.2	1165	2	Q45844	Q45844 clostridium
39	195.5	8.2	1165	2	Q45887	Q45887 clostridium
40	194	8.1	1160	2	O87710	O87710 clostridium
41	191.5	8.0	1159	2	P71109	P71109 clostridium
42	191.5	8.0	1162	2	Q9ZAJ6	Q9zaj6 clostridium
43	183.5	7.7	1196	2	Q9LBR2	Q9lbr2 clostridium
44	183	7.6	1642	5	Q8IB84	Q8ib84 plasmodium
45	182.5	7.6	441	5	Q8T1W3	Q8t1w3 dictyosteli

ALIGNMENTS

RESULT 1
Q9QTG7
ID Q9QTG7 PRELIMINARY; PRT; 1275 AA.
AC Q9QTG7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NTX (Fragment).
GN NTX.
OS Clostridium botulinum D bacteriophage.
OC Viruses.
OX NCBI_TaxID=29342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1873;
RX MEDLINE=99017546; PubMed=9802560;
RA Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT "Molecular composition of the 16S toxin produced by a Clostridium
RT botulinum type D strain, 1873.";
RL Microbiol. Immunol. 42:599-605(1998).
DR EMBL; AB012112; BAA75084.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1275 1275
SQ SEQUENCE 1275 AA; 146742 MW; 3C50F46C8233E2D6 CRC64;

```
  Query Match             92.9%;  Score 2226;  DB 12;  Length 1275;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-112;
  Matches  423;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 910

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 970

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1150

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 1210

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 1270

Qy        446 DPGWV 450
              |||||
Db       1271 DPGWV 1275
```

```
RESULT 2
Q45849
ID   Q45849      PRELIMINARY;      PRT;  1280 AA.
AC   Q45849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum C.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=36828;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6813;
RX   MEDLINE=96156810; PubMed=8593068;
RA   Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA   Oguma K.;
RT   "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT   composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT   detection of this gene from Clostridium botulinum type C organisms.";
RL   Appl. Environ. Microbiol. 62:662-667(1996).
DR   EMBL; D49440; BAA08418.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147815 MW;  74F813B228B8C989 CRC64;
```

```
  Query Match             88.9%;  Score 2131;  DB 2;  Length 1280;
  Best Local Similarity   94.1%;  Pred. No. 3.7e-107;
  Matches  401;  Conservative   18;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :|||||||||||||||||||||||||||||||||||||:  :||:||||||||||||||
Db        855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSS 914

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        915 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIKQNSGWKLCIRNGN 974

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              |||||||:||||||||||||||||||||||||||||||||||||||||||||||::||||
Db        975 IEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSERIE 1034

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 DLNEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1094

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              ||||||||||||||||||:||:||||:|||||||||||||||||||||||||||||||||
Db       1095 NPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1154

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              |||||: |||||| |||||||||||||| :| |||:|||||||||||||||||||||||||
Db       1155 NGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSKNCVYALKLQSNLGNYGIGIFSIKNI 1214

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||:|||||||||||| :|||||||||||||||:|||||||||||||||||||||||||||
Db       1215 VSQNKYCSQIFSSFMKNTMLLADIYKPWRFSFENAYTPVAVTNYETKLLSTSSFWKFISR 1274

Qy        446 DPGWVE 451
              ||||||
Db       1275 DPGWVE 1280
```

```
RESULT 3
Q9LBS7
ID   Q9LBS7      PRELIMINARY;      PRT;  1280 AA.
AC   Q9LBS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-6814;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Kawabe T.,
RA   Murakami F., Nakatsuka M., Ohyama T.;
RT   "Organization of Gene Encoding Components of the Botulinum Progenitor
RT   Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT   Chimeric Sequence in the Gene Encoding Each Component.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037166; BAA89713.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
```

```
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147757 MW;  DE124FFB6F68450B CRC64;

  Query Match             88.9%;  Score 2131;  DB 2;  Length 1280;
  Best Local Similarity   94.1%;  Pred. No. 3.7e-107;
  Matches  401;  Conservative   18;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :|||||||||||||||||||||||||||||||||||||||||:  :||:||||||||||||
Db        855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSS 914

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        915 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIKQNSGWKLCIRNGN 974

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||::||
Db        975 IEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSERIE 1034

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 DLNEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1094

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              |||||||||||||||||||||:||:||||:|||||||||||||||||||||||||||||
Db       1095 NPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1154

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              |||||: |||||| |||||||||||||| :| |||:|||||||||||||||||||||||||
Db       1155 NGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSKNCVYALKLQSNLGNYGIGIFSIKNI 1214

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||:|||||||||||| :|||||||||||||||:|||||||||||||||||||||||||||
Db       1215 VSQNKYCSQIFSSFMKNTMLLADIYKPWRFSFENAYTPVAVTNYETKLLSTSSFWKFISR 1274

Qy        446 DPGWVE 451
              ||||||
Db       1275 DPGWVE 1280


RESULT 4
Q841S3
ID   Q841S3      PRELIMINARY;      PRT;  1280 AA.
AC   Q841S3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2003;
RA   Lee M.S., Cheng M.C., Chen Y.P., Shiau J.R., Lin S.Y.;
RT   "An outbreak of botulism in black-faced spoonbill in Taiwan.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY251553; AAP06952.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1280 AA;  147642 MW;  34C1A2E573EE0AE7 CRC64;

  Query Match             88.7%;  Score 2125;  DB 2;  Length 1280;
  Best Local Similarity   93.9%;  Pred. No. 7.7e-107;
  Matches  400;  Conservative   18;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :|||||||||||||||||||||||||||||||||||||||||:  :||:||||||||||||
Db        855 SLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSS 914

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIRNGN 145
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        915 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIKQNSGWKLCIRNGN 974

Qy        146 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 205
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||::||
Db        975 IEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSERIE 1034

Qy        206 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 265
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 DLNEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1094

Qy        266 NPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKSVSDKNPYSRIL 325
              |||||||||||||||||||||:||:||||:|||||||||||||||||||||||||||||
Db       1095 NPLKFDTEYYIINDNYIDRYIAPKSNILVLVQYPDRSKLYTGNPITIKSVSDKNPYSRIL 1154

Qy        326 NGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNI 385
              |||||: |||||| |||||||||||||| :| |||:||||||| ||||||||||||||||
Db       1155 NGDNIMFHMLYNSGKYMIIRDTDTIYAIEGRECSKNCVYALKSQSNLGNYGIGIFSIKNI 1214

Qy        386 VSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISR 445
              ||:|||||||||||| :|||||||||||||||:|||||||||||||||||||||||||||
Db       1215 VSQNKYCSQIFSSFMKNTMLLADIYKPWRFSFENAYTPVAVTNYETKLLSTSSFWKFISR 1274

Qy        446 DPGWVE 451
              ||||||
Db       1275 DPGWVE 1280


RESULT 5
Q93HT3
ID   Q93HT3      PRELIMINARY;      PRT;  1291 AA.
AC   Q93HT3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-Yoichi;
RX   MEDLINE=21534265; PubMed=11676492;
RA   Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,
RA   Fujinaga Y., Oguma K., Ohyama T.;
RT   "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in
RT   Hemagglutination.";
RL   Biochem. Biophys. Res. Commun. 288:650-657(2001).
DR   EMBL; AB061780; BAB71749.1; -.
```

```
DR   PIR; A43503; A43503.
DR   PIR; S11291; A49777.
DR   PIR; S46431; S46431.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  148869 MW;  4A21DB35B8743CF8 CRC64;

  Query Match             33.4%;  Score 801.5;  DB 2;  Length 1291;
  Best Local Similarity   41.7%;  Pred. No. 2.8e-35;
  Matches  183;  Conservative   83;  Mismatches  146;  Indels   27;  Gaps   11;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :|||||||||||:|||||||||||:|| ||||||||||     :|||| |:  |||| ||
Db        855 SLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSS 914

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              |:   |:||  | ||:|:::||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        915 GEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 973

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:      :|: | |  | :  || |||||||:|||:|| ||:||||:|  :
Db        974 SNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTI 1032

Qy        203 KIEDLDEVKLDKTIVFGIDENID--------ENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| | |::  |        :|  :||||| ||:|||  :||||::
Db       1033 KVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSL 1092

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
              ||:|||||| |::: |||::| :|::||:   |  :|     : :    |  | ||
Db       1093 QYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKR 1152

Qy        315 VSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGN 374
              :      :|:  || :   |  |:: | :    :|:||      |  :|||: |:    :
Db       1153 IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH---STEDIYAIGLREQTKD 1209

Qy        375 YGIG-IFSIKNIVSKNKYCSQIF-SSFR-ENTMLLADIYKPWRFS-----FKNAYTPVAV 426
                 || |: : :   | |||| |:|  ||   :  |   :||      ::: |    |
Db       1210 INDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSI-GTYRFRLGGDWYRHNYLVPTV 1268

Qy        427 T--NYETKLLSTSSFWKFI 443
              || : | |||: | |:
Db       1269 KQGNYASLLESTSTHWGFV 1287


RESULT 6
Q45967
ID   Q45967      PRELIMINARY;      PRT;  1285 AA.
AC   Q45967;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE D;
RX   MEDLINE=96283801; PubMed=8679691;
RA   Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA   Ogumad K.;
RT   "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT   types C and D organisms.";
RL   Biochim. Biophys. Acta 1307:123-126(1996).
DR   EMBL; D38442; BAA07477.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1285 AA;  147365 MW;  EEDE98E4ECA6C413 CRC64;

  Query Match             32.6%;  Score 782.5;  DB 2;  Length 1285;
  Best Local Similarity   40.4%;  Pred. No. 3e-34;
  Matches  188;  Conservative   73;  Mismatches  123;  Indels   81;  Gaps   14;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||:||||||||||||||||||| |:||||||||||  |||||  |:  |||| ||
Db        851 SLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSS 910

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              ||   ||||  | ||:|:|:||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        911 GDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 969

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:      : : | |  | :  || |||||||| |:|| | :||||:|  :
Db        970 SNFLVFTLKQNENSEQDINFSYDISKNAAGY-NKWFFVTITTNMMGNMMIYINGKLIDTI 1028

Qy        203 KIEDLDEVKLDKTIVF--------GIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| |         |:  :  :|  :||||| ||:||| ::||||::
Db       1029 KVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSL 1088

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
              ||:|||||| |::| |||:|| ||::||:: : | :|     : :    |  | ||
Db       1089 QYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFNEGYKIIIKR 1148

Qy        315 VSDKNPYSRILNGDNII-----------LHMLYNS--------------------RKY- 341
              :      :|:  |:|::           | |  |                      |||
Db       1149 IRGNTNDTRV-RGENVLYFNTTIDNKQYSLGMYKPSRNLGTDLVPLGALDQPMDEIRKYG 1207

Qy        342 -MIIRDTDTI--YATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSS 398
               ||:  :|   ||:|           | | ||     :|| || :             |
Db       1208 SFIIQPCNTFDYYASQ-----------LFLSSNATTNRLGILSIGSY-----------S 1244

Qy        399 FRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFI 443
              |:     | | |  | |: :     : : :| : | |||: | |:
Db       1245 FK-----LGDDY--W-FNHEYLIPVIKIEHYASLLESTSTHWVFV 1281


RESULT 7
Q9LBR1
ID   Q9LBR1      PRELIMINARY;      PRT;  1285 AA.
AC   Q9LBR1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
```

```
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-4947;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA   Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT   "Characterization of the Progenitor Toxin Components Produced by
RT   Clostridium botulinum Type D Strain 4947.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037920; BAA90661.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1285 AA;  147352 MW;  B63AFA487D570680 CRC64;

  Query Match           32.6%;  Score 781.5;  DB 2;  Length 1285;
  Best Local Similarity  40.2%;  Pred. No. 3.4e-34;
  Matches  187;  Conservative   74;  Mismatches  123;  Indels   81;  Gaps   14;

Qy         26 ALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS 85
              :||||:||||||||||||||||||||| |:|||||||||| ||||| |:  |||| ||
Db        851 SLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSS 910

Qy         86 GD---KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNSGWKLCIR 142
              ||   |:||  | ||:|:|:||: |:||||:|:| ::|    ||||:|:: |||| : |
Db        911 GDDRGKVIVTQNENIVYNAMYESFSISFWIRINKWVSNLPG-YTIIDSVKNNSGWSIGII 969

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              :  : : |:      : : | |  | :  || |||||||| |:|| | :||||:|  :
Db        970 SNFLVFTLKQNENSEQDINFSYDISKNAAGY-NKWFFVTITTNMMGNMMIYINGKLIDTI 1028

Qy        203 KIEDLDEVKLDKTIVF--------GIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQ 254
              |:::|  :   ||| |        |:  :  :|  :||||| ||:||| ::||||::
Db       1029 KVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSL 1088

Qy        255 ILRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYPDRSKLYTGNPITIKS 314
                ||:|||||| |::| |||:|| ||::||:: : | :|     : :    |  | ||
Db       1089 QYTNVVKDYWGNDLRYDKEYYMINVNYMNRYMSKKGNGIVFNTRKNNNDFNEGYKIIIKR 1148

Qy        315 VSDKNPYSRILNGDNII------------LHMLYNS--------------------RKY- 341
              :    :|:  |:|::              | |   |                      |||
Db       1149 IRGNTNDTRV-RGENVLYFNTTIDNKQYSLGMYKPSRNLGTDLVPLGALDQPMDEIRKYG 1207

Qy        342 -MIIRDTDTI--YATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSS 398
               ||:  :|   ||:|           | | ||     :|| || :             |
Db       1208 SFIIQPCNTFDYYASQ-----------LFLSSNATTNRLGILSIGSY------------S 1244

Qy        399 FRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFI 443
              |:     | | |  | |: :     : : :| : | |||: | |:
Db       1245 FK-----LGDDY--W-FNHEYLIPVIKIEHYASLLESTSTHWVFV 1281


RESULT 8
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match           26.1%;  Score 626.5;  DB 2;  Length 1291;
  Best Local Similarity  34.2%;  Pred. No. 8.2e-26;
  Matches  162;  Conservative   75;  Mismatches  158;  Indels   79;  Gaps   14;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              :| :| |:| : | :: ||:|: : | |:| ||| |:| | | |:||    | |||:||
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         87 D-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
              : || |  | ||:::::: : ||||||:| |   :       |||||||| :: |||||:
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              ||   | | | |:| | ||: |:||     : | |:|||||||||     |:||||:|:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKLES 1024

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
                | |: ||  :  |:| :| |||  | :|:: |:||: |||  :|  :|: |     :
Db       1025 HIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYL 1084

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRSK------LYTGNP 309
              ||:||||| :: |||: |    : ||      :|   :|   :|   ||       || |
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEK 1144

Qy        310 ITIKSVSDKNPYSRILNGD------------------NIILHMLYNSRKYMI----IRD 346
                |:  |:    |: :| |                  : ::  :   :  :    | |
Db       1145 FIIRRKSN----SQSINDDIVRKEDYIYLDFFNLNQEWRVYIYKYFKKEEEKLFLAPISD 1200

Qy        347 TDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFS 397
              :| | | |  |  :   |: :|           :|  ||  |    ||
Db       1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV---------- 1249

Qy        398 SFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
              |:|        || :     |      |     | :||       |:|| :| || |
Db       1250 -FKE--------YKDYFCISKWYLKEVKRKPYNSKL---GCNWQFIPKDEGWTE 1291


RESULT 9
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
```

```
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match           26.1%;  Score 624.5;  DB 2;  Length 1291;
  Best Local Similarity  34.6%;  Pred. No. 1e-25;
  Matches  164;  Conservative   72;  Mismatches  159;  Indels   79;  Gaps   14;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              :| :| |:| : | :: ||:|: : | |:| ||| |:| | | |:||    | |||:||
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         87 D-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
              : || |  | ||:::::: : |||||||:| |   :      |||||||| :: |||||:
Db        906 NSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              ||   | | | |:| | ||: |:||     : | |:|||||||||     |:||||:|:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKLES 1024

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
              | |: ||  :  |:| :| |||  | :|:: |:||: |||  :|  :|: |     :
Db       1025 HIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYL 1084

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRSK------LYTGNP 309
              ||:||||| :: |||: |    : ||     :    :|  :|   ||      || |
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGEK 1144

Qy        310 ITIKSVSDKNPYSRILNGDNI---------ILHMLYNSRKYMI--------------IRD 346
              |:  |:    |: :| | :           ::    | ||               | |
Db       1145 FIIRRKSN----SQSINDDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLFLAPISD 1200

Qy        347 TDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFS 397
              :|  | | |  | :   |: :|          :|  || |    ||
Db       1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV---------- 1249

Qy        398 SFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
               |:|        || :    |      |    | :||      |:|| :| || |
Db       1250 -FKE--------YKDYFCISKWYLKEVKRKPYNSKL---GCNWQFIPKDEGWTE 1291


RESULT 10
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match           25.9%;  Score 621.5;  DB 2;  Length 1291;
  Best Local Similarity  34.6%;  Pred. No. 1.5e-25;
  Matches  164;  Conservative   72;  Mismatches  159;  Indels   79;  Gaps   14;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              :| :| |:| : | :: ||:|: : | |:| ||| |:| | | |:||    | |||:||
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         87 D-KIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
              : || |  | ||:::::: : |||||||:| |   :      |||||||| :: |||||:
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              ||   | | | |:| | ||: |:||     : | |:|||||||||     |:||||:|:
Db        966 IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNN-SDNAKIYINGKLES 1024

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
              | |: ||  :  |:| :| |||  | :|:: |:||: |||  :|  :|: |     :
Db       1025 HIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSEYL 1084

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRSK------LYTGNP 309
              ||:||||| :: |||: |    : ||     :    :|  :|   ||      || |
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIGEK 1144

Qy        310 ITIKSVSDKNPYSRILNGDNI---------ILHMLYNSRKYMI--------------IRD 346
              |:  |:    |: :| | :           ::    | ||               | |
Db       1145 FIIRRKSN----SQSINDDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLFLAPISD 1200

Qy        347 TDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYCSQIFS 397
              :|  | | |  | :   |: :|          :|  || |    ||
Db       1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIV---------- 1249

Qy        398 SFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
               |:|        || :    |      |    | :||      |:|| :| || |
Db       1250 -FKE--------YKDYFCISKWYLKEVKRKPYNSKL---GCNWQFIPKDEGWTE 1291


RESULT 11
Q8GR96
```

```
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             25.8%;  Score 617.5;  DB 2;  Length 1291;
  Best Local Similarity   33.8%;  Pred. No. 2.5e-25;
  Matches  158;  Conservative   81;  Mismatches  162;  Indels   67;  Gaps   15;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL-SSS 85
Db        848 ILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELND--KNQFKLTSST 905

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
Db        906 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKIS 965

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
Db        966 IRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNN-SDNAKIYINGKLES 1024

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
Db       1025 NIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYL 1084

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRS------KLYTGNP 309
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEK 1144

Qy        310 ITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALK-- 367
Db       1145 FIIRRKSN----SQSINDD-------------IVRKEDYIYLDFFNSNREWRVYAYKDF 1186

Qy        368 -------LQSNL--GNYGIGIFSIKNIVSKNKY-CSQIFSSFRENT--MLLADIYKPWRF 415
Db       1187 KEEEKKLFLANIYDSNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES 1246

Qy        416 -----SFKNAYTPVAVTNYETKLLSTSSF-------WKFISRDPGWVE 451
Db       1247 GIVLKDYKNYF---CISKWYLKEVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 12
```

```
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match             25.7%;  Score 615.5;  DB 2;  Length 1291;
  Best Local Similarity   33.0%;  Pred. No. 3.2e-25;
  Matches  158;  Conservative   88;  Mismatches  152;  Indels   81;  Gaps   16;

Qy         23 ASMALLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL 82
Db        844 SNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKL 901

Qy         83 SSSGD-KIIVNLNNNILYSAIYENSSVSFWIKISK----DLTN-SHNEYTIINSIEQNSG 136
Db        902 TSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSG 961

Qy        137 WKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYING 196
Db        962 WKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYING 1020

Qy        197 ELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQIL 256
Db       1021 TLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSY 1080

Qy        257 RNVIKDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRS------KLY 305
Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLY 1140

Qy        306 TGNPITIKSVSDKNPYSRILNGD----------NIILH----MLYNSRKYM--------- 342
Db       1141 IGEKFIIRRESN----SQSINDDIVRKEDYIHLDLVLHHEEWRVY-AYKYFKEQEEKLFL 1195

Qy        343 -IIRDTDTIYAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIVSKNKYC 392
Db       1196 SIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKK--- 1252
```

```
Qy        393 SQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
                                 || :    |      |    |:: |      |:|| :| || |
Db       1253 ----------------YKDYFCISKWYLKEVKRKPYKSNL---GCNWQFIPKDEGWTE 1291


RESULT 13
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;

  Query Match             24.7%;  Score 592.5;  DB 2;  Length 441;
  Best Local Similarity   35.3%;  Pred. No. 1.6e-24;
  Matches  146;  Conservative   72;  Mismatches  145;  Indels   51;  Gaps   11;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL-SSS 85
              :| :| |:| : | :: ||:|: : | |:| ||| | | | | |:||    | ||| ||:
Db          6 ILIEIFNKYNSKILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELND--KNQFKLTSST 63

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWKLC 140
                :| |  | ||:::::: : ||||||:| |   :       |||||||| |: |||||:
Db         64 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWKIS 123

Qy        141 IRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQ 200
              ||   | | | |:| | ||: |:||     : | |:||||||||      |:||||:|:
Db        124 IRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNN-SDNAKIYINGKLES 182

Qy        201 SQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVI 260
              :  |:|: ||  :  |:| :| :||  | :|:: |:||: |||  :|  :|: |    :
Db        183 NIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSEYL 242

Qy        261 KDYWGNPLKFDTEYYIINDNYIDRYI-----APESNVLVLVRYPDRS------KLYTGNP 309
              ||:|||| :: |||: |    : ||      :   :|   :|   |        || |
Db        243 KDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIGEK 302

Qy        310 ITIKSVSDKNPYSR--ILNGDNIILHMLYNSRKYMI-----------------IRDTDTI 350
              |:  |:    :   :  | | |    ::|:: :                   | |::
Db        303 FIIRRKSNSQSINDDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLVLANIYDSNEF 362

Qy        351 YAT-QGGECSQNCVYALKL--------QSNLGNYGIGIFSIKNIV---SKNKYC 392
              | | |  |  :  |: :|          :|  ||  |    ||   || :|
Db        363 YKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFC 416


RESULT 14
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match             24.0%;  Score 575.5;  DB 2;  Length 1268;
  Best Local Similarity   28.5%;  Pred. No. 4.6e-23;
  Matches  133;  Conservative   95;  Mismatches  155;  Indels   83;  Gaps   12;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG 86
              :|  |:  ::  | || ||::: : |  :|:||| : : :  :: : :   -| | : ||
Db        845 ILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYSSR 904

Qy         87 -DKIIVNLNNNILYSAIYENSSVSFWIKISK--DLTNSHNEYTIINSI-EQNSGWKLCIR 142
               :: :  || |:|:: |:| |||||::|  |  :| | :|||||||  :   |||||: :
Db        905 LSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNNLKNLNNEYTIINCMRNNNSGWKISLN 964

Qy        143 NGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQ 202
              || | |||      : |:|:|:: :  :  | ||| ||||||| :|: ||||||  |   :
Db        965 YNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGNLTDQK 1024

Qy        203 KIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKD 262
              | :|  : :|  |:| |    :: : : || | ||: ||   :|  :|   :  ::||
Db       1025 SILNLGNIHVDDNILFKI-VGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDSTILKD 1083

Qy        263 YWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVRYP---------DRSKLYTGNPITIK 313
              :||| | :: :||::|    :  :   |::| : |                ::|||  : |:
Db       1084 FWGNYLLYNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNIFSNARLYTGVEVIIR 1143

Qy        314 SVSD---KNPYSRILNGDNIILHMLYNSRKYMIIRDTDT---------------------- 349
              |      |  : :   | : :::: : :| : |  | |
Db       1144 KVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRISNSNYNSNQ 1203

Qy        350 --IYATQGGECSQNCVYALKLQSNLGNYGIGI--FSIKNIVSKNKYCSQIFSSFRENTML 405
                | : |  |: |       ::| ||  ||:  | : |:|: : |    : : | ||
Db       1204 MIIMDSIGDNCTMN------FKTNNGN-DIGLLGFHLNNLVASSWY----YKNIRNNTR- 1251

Qy        406 LADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 451
                                      :   || |||:: || |
Db       1252 ----------------------------NNGCFWSFISKEHGWQE 1268
```

```
RESULT 15
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             23.8%;  Score 571.5;  DB 2;  Length 1251;
  Best Local Similarity   31.4%;  Pred. No. 7.4e-23;
  Matches  142;  Conservative   81;  Mismatches  166;  Indels   63;  Gaps   14;

Qy         27 LLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKL-SSS 85
              :|    |::|  |  | :|::: | :  ||||||:: : :   : :     | | : :|
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFNSK 894

Qy         86 GDKIIVNLNNNILYSAIYENSSVSFWIKI---SKDLTNSHNEYTIINSI-EQNSGWKLCI 141
                :: :: |: |:|   |:| |:|||::|      : | :|||||| : : |||||: :
Db        895 PSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNINNEYTIINCMRDNNSGWKVSL 954

Qy        142 RNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQS 201
              :  | | |||  |  : |:| |  :   : | ||| ||||||: :|  |||||| |
Db        955 NHNEIIWTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGHLIDQ 1014

Qy        202 QKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIK 261
              : | :|  : :   |:| | |    : : || |||| |||   :|  :|  :   |::|
Db       1015 KSILNLGNIHVSDNILFKI-VNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILK 1073

Qy        262 DYWGNPLKFDTEYYIIN----DNYIDRYIAP-------ESNVLVLVRYPDRSKLYTGNPI 310
              |:|| | :|  ||::|    :|:|||           | :|:       ::||:|  :
Db       1074 DFWGNYLLYDKGYYLLNVLKPNNFIDRRKDSTLSINNIRSTILL------ANRLYSGIKV 1127

Qy        311 TIKSVSDKNPYSR-ILNGDNIILHMLYNSRKYMIIRDTD------TIYATQGGECSQNCV 363
               |: |:| :   | :   | : :: : ||  | :  ||:      || ::  |      |
```

```
Db       1128 KIQRVNDSSTNDRFVRKNDQVYINYISNSSSYSLYADTNTTDKEKTIKSSSSGNRFNQVV 1187

Qy        364 YALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTP 423
                              : | |  |  |:  | :   | : |             || | |
Db       1188 ------------------VMNSVGNN--CTMNFKNNNGNNIGL--------LGFK-ADTV 1218

Qy        424 VAVTNYETKLL----STSSFWKFISRDPGWVE 451
              || | | | :     |  ||| ||| : || |
Db       1219 VASTWYYTHMRDHTNSNGCFWNFISEEHGWQE 1250
```

Search completed: September 10, 2004, 03:41:46
Job time : 22.6355 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 26.2491 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 448
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	448	100.0	448	2	AAW68399	Aaw68399 Clostridi
2	27	6.0	144	2	AAW09016	Aaw09016 Immunogen
3	27	6.0	431	2	AAW09014	Aaw09014 Immunogen
4	27	6.0	432	3	AAY77138	Aay77138 Synthetic
5	27	6.0	432	4	AAB04103	Aab04103 Botulism
6	27	6.0	432	4	AAB04096	Aab04096 Botulism
7	27	6.0	645	4	AAE07894	Aae07894 Modified
8	27	6.0	645	6	AAE35692	Aae35692 DipT HN d
9	27	6.0	657	6	AAE35693	Aae35693 BoNT/F-Hc
10	27	6.0	657	6	AAE35694	Aae35694 BoNT/F-Hc
11	27	6.0	660	4	AAE07898	Aae07898 Modified
12	27	6.0	685	4	AAE07893	Aae07893 Modified
13	27	6.0	862	4	AAE07890	Aae07890 Modified
14	27	6.0	887	4	AAE07892	Aae07892 Modified
15	27	6.0	979	6	AAE35713	Aae35713 BoNT/F-Hc
16	27	6.0	1032	4	AAE07901	Aae07901 C. botuli
17	27	6.0	1059	3	AAY93309	Aay93309 A mangane
18	27	6.0	1084	3	AAY93312	Aay93312 A mangane
19	27	6.0	1092	4	AAE07900	Aae07900 C. botuli
20	27	6.0	1192	6	AAE35711	Aae35711 BoNT/F-Hc
21	27	6.0	1192	6	AAE35710	Aae35710 BoNT/F-Hc
22	26	5.8	144	2	AAW09015	Aaw09015 Immunogen
23	26	5.8	451	2	AAW68395	Aaw68395 Clostridi
24	26	5.8	451	2	AAW68398	Aaw68398 Clostridi
25	26	5.8	452	2	AAW68396	Aaw68396 Clostridi
26	26	5.8	462	2	AAR95009	Aar95009 Type A ne
27	26	5.8	462	2	AAW68390	Aaw68390 Clostridi
28	26	5.8	462	2	AAW68397	Aaw68397 Clostridi
29	26	5.8	472	2	AAW68393	Aaw68393 Clostridi
30	26	5.8	472	2	AAW68394	Aaw68394 Clostridi
31	26	5.8	473	2	AAW68400	Aaw68400 Clostridi
32	24	5.4	143	2	AAW09017	Aaw09017 Immunogen
33	23	5.1	393	4	AAG68149	Aag68149 Streptomy
34	23	5.1	393	4	AAB74844	Aab74844 S. griseu
35	23	5.1	423	2	AAW68098	Aaw68098 Chlamydom
36	22	4.9	22	4	AAB67674	Aab67674 Amino aci
37	22	4.9	22	5	AAM49518	Aam49518 His-tag p
38	22	4.9	26	7	ADD31745	Add31745 Purificat
39	22	4.9	53	2	AAW36986	Aaw36986 HFKT pept
40	22	4.9	54	2	AAW36990	Aaw36990 HFKT pept
41	22	4.9	91	2	AAY09127	Aay09127 Recombina
42	22	4.9	93	2	AAY09128	Aay09128 Recombina
43	22	4.9	195	3	AAB11377	Aab11377 B. stearo
44	22	4.9	269	5	AAM49520	Aam49520 B. mori 1
45	22	4.9	269	5	AAM49541	Aam49541 B. mori T

ALIGNMENTS

RESULT 1
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30593.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC express BotC soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which

CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX
SQ Sequence 448 AA;

Query Match 100.0%; Score 448; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
```

RESULT 2
AAW09016
ID AAW09016 standard; protein; 144 AA.
XX
AC AAW09016;
XX
DT 17-OCT-2003 (revised)
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin polypeptide (aa992-1135).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum; type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB001409.
XX
PR 12-JUN-1995; 95GB-00011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT vaccine prodn.
XX
PS Claim 5; Page 18-19; 37pp; English.
XX
CC Novel polypeptides (AAW09014-17) respectively comprise amino acids 848-
CC 1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a type F
CC botulinum neurotoxin (BoNT/F). They lack the L chain and HN epitopes
CC necessary for metalloprotease activity and toxin internalisation. They
CC are free of botulinum toxin activity but can induce protective immunity
CC to a type F botulinum toxin, making them useful for vaccine prodn.
CC Recombinant polypeptides can be produced in transformed host cells, esp.
CC as fusion proteins, e.g. with maltose binding protein to facilitate
CC purification. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 144 AA;

Query Match 6.0%; Score 27; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db         13 YINKWIFVTITNNRLGNSRIYINGNLI 39
```

RESULT 3
AAW09014
ID AAW09014 standard; protein; 431 AA.
XX
AC AAW09014;
XX
DT 17-OCT-2003 (revised)
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum; type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB001409.
XX
PR 12-JUN-1995; 95GB-00011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX
CC A polypeptide (AAW09014) comprises the heavy chain (amino acids 848-1278)
CC of a type F botulinum neurotoxin (BoNT/F), and can be produced using a
CC synthetic gene (AAT48101) based on the natural gene sequence (AAT48100)
CC for the heavy chain. The polypeptides and its fragments (see also
CC AAW09015-17) lack the light chain and HN epitopes necessary for
CC metalloprotease activity and toxin internalisation. They are free of
CC botulinum toxin activity but can induce protective immunity to a type F
CC botulinum toxin, making them useful for vaccine prodn. Recombinant
CC polypeptides can be produced in transformed host cells, esp. as fusion
CC proteins, e.g. with maltose binding protein to facilitate purification.

CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 431 AA;

Query Match 6.0%; Score 27; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.5e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 YINKWIFVTITNNRLGNSRIYINGNLI 206
||||||||||||||||||||||||||||
Db 157 YINKWIFVTITNNRLGNSRIYINGNLI 183

RESULT 4
AAY77138
ID AAY77138 standard; protein; 432 AA.
XX
AC AAY77138;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype F; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87216.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 27; Page; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast. Note: This
CC sequence is not given in the specification, but is decoded from the BoNTF
CC Hc DNA sequence given on pages 45-46
XX
SQ Sequence 432 AA;

Query Match 6.0%; Score 27; DB 3; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.5e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 YINKWIFVTITNNRLGNSRIYINGNLI 206
||||||||||||||||||||||||||||
Db 158 YINKWIFVTITNNRLGNSRIYINGNLI 184

RESULT 5
AAB04103
ID AAB04103 standard; protein; 432 AA.
XX
AC AAB04103;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54499.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Fig 18b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G

```
XX
SQ   Sequence 432 AA;

  Query Match             6.0%;  Score 27;  DB 4;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        158 YINKWIFVTITNNRLGNSRIYINGNLI 184


RESULT 6
AAB04096
ID   AAB04096 standard; protein; 432 AA.
XX
AC   AAB04096;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54490.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 9b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 432 AA;

  Query Match             6.0%;  Score 27;  DB 4;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        158 YINKWIFVTITNNRLGNSRIYINGNLI 184


RESULT 7
AAE07894
ID   AAE07894 standard; protein; 645 AA.
XX
AC   AAE07894;
XX
DT   11-SEP-2003  (revised)
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;   99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 2; Page 44; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain fragment. This sequence is
CC   constructed by fusing the binding domain of botulinum neurotoxin type F
CC   (BoNT/F) with translocation domain of diphtheria neurotoxin. (Updated on
CC   11-SEP-2003 to standardise OS field)
XX
SQ   Sequence 645 AA;

  Query Match             6.0%;  Score 27;  DB 4;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        371 YINKWIFVTITNNRLGNSRIYINGNLI 397
```

RESULT 8
AAE35692
ID AAE35692 standard; protein; 645 AA.
XX
AC AAE35692;
XX
DT 23-OCT-2003 (revised)
DT 17-JUN-2003 (first entry)
XX
DE DipT HN domain-BoNT/F-Hc fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW translocation domain; HN domain; DipT; Hc; botulinum type F neurotoxin;
KW binding domain; BoNT/F.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 57-60; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (Dip-HN domain) and botulinum type F neurotoxin from Clostridium
CC botulinum. This sequence is used in the exemplification of the invention.
CC (Updated on 23-OCT-2003 to standardise OS field)
XX
SQ Sequence 645 AA;

```
  Query Match             6.0%;  Score 27;  DB 6;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        371 YINKWIFVTITNNRLGNSRIYINGNLI 397
```

RESULT 9
AAE35693
ID AAE35693 standard; protein; 657 AA.
XX
AC AAE35693;
XX
DT 17-JUN-2003 (first entry)
XX
DE BoNT/F-Hc-DipT HN domain-thrombin linker fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW botulinum type F neurotoxin.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Unidentified.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 60-63; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and thrombin linker peptide. This
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 657 AA;

```
  Query Match             6.0%;  Score 27;  DB 6;  Length 657;
```

```
  Best Local Similarity   100.0%;  Pred. No. 6.6e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        383 YINKWIFVTITNNRLGNSRIYINGNLI 409
```

```
RESULT 10
AAE35694
ID   AAE35694 standard; protein; 657 AA.
XX
AC   AAE35694;
XX
DT   17-JUN-2003  (first entry)
XX
DE   BoNT/F-Hc-DipT HN domain-factor Xa linker fusion construct.
XX
KW   Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW   infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW   muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW   torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW   BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW   botulinum type F neurotoxin.
XX
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Unidentified.
OS   Chimeric.
XX
PN   WO200296467-A2.
XX
PD   05-DEC-2002.
XX
PF   21-MAY-2002; 2002WO-GB002384.
XX
PR   24-MAY-2001; 2001GB-00012687.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Sutton JM,  Shone CC;
XX
DR   WPI; 2003-167247/16.
XX
PT   Conjugate for modulating cell survival and cell growth, modulating
PT   release of inflammatory mediator from cells, comprises injected bacterial
PT   effector protein and a carrier that targets the protein to target cell.
XX
PS   Example 12; Page 63-65; 130pp; English.
XX
CC   The invention relates to a conjugate comprising an injected bacterial
CC   effector protein and a carrier that targets the effector protein to a
CC   target cell. Pharmaceutical composition of the invention is useful for a
CC   treatment selected from promoting or inhibiting survival of cells;
CC   preventing and reversing damage to cells; killing cells; promoting or
CC   inhibiting the growth of cells, apoptosis, release of an inflammatory
CC   mediator from cells, division of cells and treating intracellular
CC   infection and regulating nitric oxide release from cells. The invention
CC   is useful in the manufacture of a medicament for treating a neuronal
CC   cell, for intracellular infection, for interfering with intracellular
CC   trafficking, for modulating expression of cell-surface markers and for
CC   inhibiting secretion from cells. The invention is also useful for
CC   treating Prion disease, Alzheimer' disease and wide range of disorders
CC   including muscle spasms such as blephorospasm, torticolis and
CC   hypersecretion disorders such as chronic obstructive pulmonary disease
CC   (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC   comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC   domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC   (BoNT/F-Hc) from Clostridium botulinum and factor Xa linker peptide. This
CC   sequence is used in the exemplification of the invention
XX
SQ   Sequence 657 AA;
```

```
  Query Match             6.0%;  Score 27;  DB 6;  Length 657;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        383 YINKWIFVTITNNRLGNSRIYINGNLI 409
```

```
RESULT 11
AAE07898
ID   AAE07898 standard; protein; 660 AA.
XX
AC   AAE07898;
XX
DT   11-SEP-2003  (revised)
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #5.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; tetanus neurotoxin; TeNT;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Corynebacterium diphtheriae.
OS   Clostridium tetani.
OS   Clostridium botulinum.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;   99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 2; Page 46; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain fragment. This sequence is
CC   constructed by fusing the binding domain which is a hybrid of botulinum
CC   neurotoxin type F (BoNT/F) and tetanus neurotoxin (TeNT) domain II with
CC   translocation domain of diphtheria neurotoxin. (Updated on 11-SEP-2003 to
CC   standardise OS field)
XX
SQ   Sequence 660 AA;
```

```
  Query Match             6.0%;  Score 27;  DB 4;  Length 660;
```

```
  Best Local Similarity   100.0%;  Pred. No. 6.7e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        371 YINKWIFVTITNNRLGNSRIYINGNLI 397


RESULT 12
AAE07893
ID   AAE07893 standard; protein; 685 AA.
XX
AC   AAE07893;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS   Geobacillus stearothermophilus.
OS   Influenza virus.
OS   Clostridium botulinum.
OS   Synthetic.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;   99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 9; Page 43; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC   conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC   Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC   translocation peptide from influenza virus and a neuronal cell-specific
CC   binding domain from botulinum neurotoxin type F (BoNT/F)
XX
SQ   Sequence 685 AA;

  Query Match             6.0%;  Score 27;  DB 4;  Length 685;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        411 YINKWIFVTITNNRLGNSRIYINGNLI 437


RESULT 13
AAE07890
ID   AAE07890 standard; protein; 862 AA.
XX
AC   AAE07890;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Geobacillus stearothermophilus.
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Synthetic.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;   99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 9; Page 40; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC   conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC   Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC   translocation domain from diphtheria neurotoxin and a neuronal cell-
CC   specific binding domain from botulinum neurotoxin type F (BoNT/F)
XX
SQ   Sequence 862 AA;

  Query Match             6.0%;  Score 27;  DB 4;  Length 862;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        588 YINKWIFVTITNNRLGNSRIYINGNLI 614
```

RESULT 14
AAE07892
ID AAE07892 standard; protein; 887 AA.
XX
AC AAE07892;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Homo sapiens.
OS Geobacillus stearothermophilus.
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Synthetic.
OS Chimeric.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB004644.
XX
PR 02-DEC-1999; 99GB-00028530.
PR 07-APR-2000; 2000GB-00008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that translocates
PT the therapeutic agent into the neuronal cells.
XX
PS Example 9; Page 42; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC conjugate comprises a mitochondrial leader sequence from human Mn-
CC superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC linker that can be cleaved by thrombin, translocation domain from
CC diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC botulinum neurotoxin type F (BoNT/F)
XX
SQ Sequence 887 AA;

```
  Query Match             6.0%;  Score 27;  DB 4;  Length 887;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        613 YINKWIFVTITNNRLGNSRIYINGNLI 639
```

RESULT 15
AAE35713
ID AAE35713 standard; protein; 979 AA.
XX
AC AAE35713;
XX
DT 17-JUN-2003 (first entry)
XX
DE BoNT/F-Hc-DipT HN domain-factor Xa linker-YopT protein fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW botulinum type F neurotoxin; targetted effector protien; YopT.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Yersinia pestis.
OS Unidentified.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 110-114; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and factor Xa linker peptide and
CC Yersinia pestis targetted effector protien YopT. This sequence is used in
CC the exemplification of the invention
XX
SQ Sequence 979 AA;

Query Match 6.0%; Score 27; DB 6; Length 979;
Best Local Similarity 100.0%; Pred. No. 9.6e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        705 YINKWIFVTITNNRLGNSRIYINGNLI 731
```

Search completed: September 10, 2004, 05:06:02
Job time : 27.2491 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 7.3452 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 448
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	5.8	462	1	US-08-480-604A-26	Sequence 26, Appl
2	26	5.8	462	2	US-08-405-496A-26	Sequence 26, Appl
3	26	5.8	462	3	US-08-915-136-26	Sequence 26, Appl
4	26	5.8	462	4	US-09-084-517-26	Sequence 26, Appl
5	23	5.1	23	1	US-08-480-604A-24	Sequence 24, Appl
6	23	5.1	23	2	US-08-405-496A-24	Sequence 24, Appl
7	23	5.1	23	3	US-08-915-136-24	Sequence 24, Appl
8	23	5.1	23	4	US-09-084-517-24	Sequence 24, Appl
9	22	4.9	22	4	US-09-660-742-3	Sequence 3, Appli
10	22	4.9	53	2	US-08-651-818A-19	Sequence 19, Appl
11	22	4.9	53	3	US-09-184-826-19	Sequence 19, Appl
12	22	4.9	54	2	US-08-651-818A-23	Sequence 23, Appl
13	22	4.9	54	3	US-09-184-826-23	Sequence 23, Appl
14	22	4.9	323	1	US-08-591-989-5	Sequence 5, Appli
15	21	4.7	21	2	US-08-651-818A-21	Sequence 21, Appl
16	21	4.7	21	3	US-09-184-826-21	Sequence 21, Appl
17	17	3.8	24	3	US-08-584-031-8	Sequence 8, Appli
18	17	3.8	24	3	US-08-780-496-8	Sequence 8, Appli
19	17	3.8	24	4	US-08-928-069-9	Sequence 9, Appli
20	15	3.3	20	4	US-09-186-489-6	Sequence 6, Appli
21	15	3.3	20	4	US-10-043-665B-6	Sequence 6, Appli
22	12	2.7	16	5	PCT-US94-04361-56	Sequence 56, Appl
23	12	2.7	17	2	US-08-600-783-13	Sequence 13, Appl
24	11	2.5	20	4	US-09-674-677-34	Sequence 34, Appl
25	11	2.5	24	4	US-09-292-858B-27	Sequence 27, Appl
26	11	2.5	254	3	US-09-167-434-7	Sequence 7, Appli
27	11	2.5	254	3	US-08-853-755-7	Sequence 7, Appli
28	11	2.5	302	2	US-08-203-532F-4	Sequence 4, Appli
29	11	2.5	302	3	US-08-950-860-16	Sequence 16, Appl
30	11	2.5	302	3	US-09-078-465-4	Sequence 4, Appli
31	11	2.5	302	5	PCT-US95-01882A-4	Sequence 4, Appli
32	11	2.5	303	2	US-08-203-532F-2	Sequence 2, Appli
33	11	2.5	303	3	US-09-078-465-2	Sequence 2, Appli
34	11	2.5	303	5	PCT-US95-01882A-2	Sequence 2, Appli
35	11	2.5	363	4	US-09-328-352-4930	Sequence 4930, Ap
36	11	2.5	414	5	PCT-US92-06840-2	Sequence 2, Appli
37	11	2.5	438	1	US-08-480-604A-23	Sequence 23, Appl
38	11	2.5	438	2	US-08-405-496A-23	Sequence 23, Appl
39	11	2.5	438	3	US-08-915-136-23	Sequence 23, Appl
40	11	2.5	438	4	US-09-084-517-23	Sequence 23, Appl
41	11	2.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
42	11	2.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
43	11	2.5	1296	3	US-08-915-136-28	Sequence 28, Appl
44	11	2.5	1296	4	US-09-084-517-28	Sequence 28, Appl
45	10	2.2	25	4	US-09-721-154-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-26

Query Match 5.8%; Score 26; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-26

Query Match 5.8%; Score 26; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26

RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-26

Query Match 5.8%; Score 26; DB 3; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;

```
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              |||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             5.8%;  Score 26;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              |||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 5
US-08-480-604A-24
; Sequence 24, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-24

  Query Match             5.1%;  Score 23;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-14;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 6
US-08-405-496A-24
```

```
; Sequence 24, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-24

  Query Match             5.1%;  Score 23;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-14;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 7
US-08-915-136-24
; Sequence 24, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-24

  Query Match             5.1%;  Score 23;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-14;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 8
US-09-084-517-24
; Sequence 24, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
```

```
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,517
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CARROLL, PETER G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  OPHD-01610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  23 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-084-517-24

  Query Match             5.1%;  Score 23;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-14;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 9
US-09-660-742-3
; Sequence 3, Application US/09660742
; Patent No. 6485977
; GENERAL INFORMATION:
;  APPLICANT:  Collmer, Alan
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  RECOMBINANT CONSTRUCTS AND TECHNIQUES FOR DELIVERING TO
;  TITLE OF INVENTION:  EUCARYOTIC CELLS BACTERIAL PROTEINS THAT ARE SECRETED
;  TITLE OF INVENTION:  VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE:  19603/4021
;  CURRENT APPLICATION NUMBER: US/09/660,742
;  CURRENT FILING DATE:  2000-09-13
;  PRIOR APPLICATION NUMBER:  60/153,507
;  PRIOR FILING DATE:  1999-09-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 22
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Polypeptide
US-09-660-742-3

  Query Match             4.9%;  Score 22;  DB 4;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 10
US-08-651-818A-19
; Sequence 19, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-19

  Query Match             4.9%;  Score 22;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 11
US-09-184-826-19
; Sequence 19, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
```

```
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-19

  Query Match             4.9%;  Score 22;  DB 3;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 12
US-08-651-818A-23
; Sequence 23, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-23

  Query Match             4.9%;  Score 22;  DB 2;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 13
US-09-184-826-23
; Sequence 23, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-23

  Query Match             4.9%;  Score 22;  DB 3;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 14
US-08-591-989-5
```

```
; Sequence 5, Application US/08591989
; Patent No. 5795721
;  GENERAL INFORMATION:
;    APPLICANT:  Ross S. Rabin, Sumedha Jayasena
;    APPLICANT:  and Larry Gold
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC
;    TITLE OF INVENTION:  ACID LIGANDS OF ICP4
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 East Prentice Avenue, Suite #200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.40 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX 49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-591-989-5

  Query Match             4.9%;  Score 22;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-12;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db         17 MGHHHHHHHHHHSSGHIEGRHM 38


RESULT 15
US-08-651-818A-21
; Sequence 21, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-21

  Query Match             4.7%;  Score 21;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-13;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRH 21
              |||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRH 21


Search completed: September 10, 2004, 05:17:09
Job time : 7.3452 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 22.6619 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 448
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	448	100.0	448	12	US-10-272-898-73	Sequence 73, Appl
2	448	100.0	448	15	US-10-354-774-73	Sequence 73, Appl
3	448	100.0	448	15	US-10-271-012-73	Sequence 73, Appl
4	448	100.0	448	16	US-10-729-122-73	Sequence 73, Appl
5	448	100.0	448	16	US-10-729-039-73	Sequence 73, Appl
6	422	94.2	1274	12	US-10-272-898-71	Sequence 71, Appl
7	422	94.2	1274	15	US-10-354-774-71	Sequence 71, Appl
8	422	94.2	1274	15	US-10-271-012-71	Sequence 71, Appl
9	422	94.2	1274	15	US-10-452-024-6	Sequence 6, Appli
10	422	94.2	1274	16	US-10-729-122-71	Sequence 71, Appl
11	422	94.2	1274	16	US-10-729-039-71	Sequence 71, Appl
12	150	33.5	369	15	US-10-452-024-174	Sequence 174, App
13	88	19.6	1280	15	US-10-452-024-162	Sequence 162, App
14	27	6.0	144	8	US-08-981-087A-3	Sequence 3, Appli
15	27	6.0	431	8	US-08-981-087A-1	Sequence 1, Appli
16	27	6.0	432	10	US-09-910-186A-16	Sequence 16, Appl
17	27	6.0	432	10	US-09-910-186A-34	Sequence 34, Appl
18	27	6.0	432	15	US-10-452-024-178	Sequence 178, App
19	27	6.0	645	14	US-10-130-973A-8	Sequence 8, Appli
20	27	6.0	660	14	US-10-130-973A-12	Sequence 12, Appl
21	27	6.0	685	14	US-10-130-973A-7	Sequence 7, Appli
22	27	6.0	862	14	US-10-130-973A-4	Sequence 4, Appli
23	27	6.0	887	14	US-10-130-973A-6	Sequence 6, Appli
24	27	6.0	1032	14	US-10-130-973A-15	Sequence 15, Appl
25	27	6.0	1092	14	US-10-130-973A-14	Sequence 14, Appl
26	27	6.0	1278	15	US-10-452-024-152	Sequence 152, App
27	27	6.0	1278	15	US-10-205-516-12	Sequence 12, Appl
28	27	6.0	1288	15	US-10-205-516-26	Sequence 26, Appl
29	26	5.8	144	8	US-08-981-087A-2	Sequence 2, Appli
30	26	5.8	366	15	US-10-452-024-173	Sequence 173, App
31	26	5.8	451	12	US-10-272-898-54	Sequence 54, Appl
32	26	5.8	451	12	US-10-272-898-68	Sequence 68, Appl
33	26	5.8	451	15	US-10-354-774-54	Sequence 54, Appl
34	26	5.8	451	15	US-10-354-774-68	Sequence 68, Appl
35	26	5.8	451	15	US-10-271-012-54	Sequence 54, Appl
36	26	5.8	451	15	US-10-271-012-68	Sequence 68, Appl
37	26	5.8	451	16	US-10-729-122-54	Sequence 54, Appl
38	26	5.8	451	16	US-10-729-122-68	Sequence 68, Appl
39	26	5.8	451	16	US-10-729-039-54	Sequence 54, Appl
40	26	5.8	451	16	US-10-729-039-68	Sequence 68, Appl
41	26	5.8	452	12	US-10-272-898-56	Sequence 56, Appl
42	26	5.8	452	15	US-10-354-774-56	Sequence 56, Appl
43	26	5.8	452	15	US-10-271-012-56	Sequence 56, Appl
44	26	5.8	452	16	US-10-729-122-56	Sequence 56, Appl
45	26	5.8	452	16	US-10-729-039-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-73
; Sequence 73, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-272-898-73

  Query Match             100.0%;  Score 448;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 2
US-10-354-774-73
; Sequence 73, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-354-774-73

  Query Match             100.0%;  Score 448;  DB 15;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 3
US-10-271-012-73
; Sequence 73, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-271-012-73

  Query Match             100.0%;  Score 448;  DB 15;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 4
US-10-729-122-73
; Sequence 73, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-729-122-73

  Query Match             100.0%;  Score 448;  DB 16;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 5
US-10-729-039-73
```

```
; Sequence 73, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-729-039-73

  Query Match           100.0%;  Score 448;  DB 16;  Length 448;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 6
US-10-272-898-71
; Sequence 71, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1274 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-272-898-71

  Query Match           94.2%;  Score 422;  DB 12;  Length 1274;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
                 ||
Db       1273 KE 1274


RESULT 7
US-10-354-774-71
; Sequence 71, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1274 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-354-774-71

  Query Match             94.2%;  Score 422;  DB 15;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
                 ||
Db       1273 KE 1274


RESULT 8
US-10-271-012-71
; Sequence 71, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1274 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-271-012-71

  Query Match             94.2%;  Score 422;  DB 15;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
              ||
Db       1273 KE 1274
```

RESULT 9
US-10-452-024-6
; Sequence 6, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-6

Query Match 94.2%; Score 422; DB 15; Length 1274;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
              ||
Db       1273 KE 1274
```

RESULT 10
US-10-729-122-71
; Sequence 71, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-729-122-71

Query Match 94.2%; Score 422; DB 16; Length 1274;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912
```

```
Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy         387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy         447 KE 448
               ||
Db        1273 KE 1274
```

RESULT 11
US-10-729-039-71
; Sequence 71, Application US/10729039
; Publication No. US20040142455A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,039
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 1996/08/28
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-729-039-71

```
  Query Match           94.2%;  Score 422;  DB 16;  Length 1274;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy         387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy         447 KE 448
               ||
Db        1273 KE 1274
```

RESULT 12
US-10-452-024-174
; Sequence 174, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 174
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-174

```
  Query Match           33.5%;  Score 150;  DB 15;  Length 369;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-132;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         220 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 279

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         280 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 339

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNR 176
```

```
                    ||||||||||||||||||||||||||||||
Db        340 TVRDCEIIWTLQDTSGNKENLIFRYEELNR 369


RESULT 13
US-10-452-024-162
; Sequence 162, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 162
;   LENGTH: 1280
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-162

  Query Match             19.6%;  Score 88;  DB 15;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-73;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIR 239
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 YINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIR 1066

Qy        240 YFKVFNTELDKTEIETLYSNEPDPSILK 267
                ||||||||||||||||||||||||||||
Db       1067 YFKVFNTELDKTEIETLYSNEPDPSILK 1094


RESULT 14
US-08-981-087A-3
; Sequence 3, Application US/08981087A
; Publication No. US20020081304A1
;  GENERAL INFORMATION:
;    APPLICANT:  Elmore, Michael J.
;    APPLICANT:  Mauchline, Margaret L.
;    APPLICANT:  Minton, Nigel P.
;    APPLICANT:  Pasechnik, Vladimir A.
;    APPLICANT:  Titball, Richard W.
;    TITLE OF INVENTION:  TYPE F BOTULINUM TOXIN AND USE THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. US20020081304A1th Glebe Rd.  8th floor
;      CITY:  Arlington
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201-4741
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,087A
;      FILING DATE:  27-MAY-1998
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/01409
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9511909.5
;      FILING DATE:  12-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford, Arthur R.
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  124-688
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  144 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-981-087A-3

  Query Match             6.0%;  Score 27;  DB 8;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-17;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
                ||||||||||||||||||||||||||||
Db         13 YINKWIFVTITNNRLGNSRIYINGNLI 39


RESULT 15
US-08-981-087A-1
; Sequence 1, Application US/08981087A
; Publication No. US20020081304A1
;  GENERAL INFORMATION:
;    APPLICANT:  Elmore, Michael J.
;    APPLICANT:  Mauchline, Margaret L.
;    APPLICANT:  Minton, Nigel P.
;    APPLICANT:  Pasechnik, Vladimir A.
;    APPLICANT:  Titball, Richard W.
;    TITLE OF INVENTION:  TYPE F BOTULINUM TOXIN AND USE THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. US20020081304A1th Glebe Rd.  8th floor
;      CITY:  Arlington
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201-4741
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,087A
;      FILING DATE:  27-MAY-1998
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/01409
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9511909.5
;      FILING DATE:  12-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford, Arthur R.
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  124-688
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  431 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
```

US-08-981-087A-1

Query Match 6.0%; Score 27; DB 8; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.1e-16;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db        157 YINKWIFVTITNNRLGNSRIYINGNLI 183
```

Search completed: September 10, 2004, 05:52:59
Job time : 23.6619 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 6.20641 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 448
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	422	94.2	1274	2	I40813	neurotoxin type F
2	150	33.5	369	2	S48109	neurotoxin type F
3	26	5.8	366	2	S48110	neurotoxin type F
4	13	2.9	1268	2	S33411	botulinum neurotox
5	12	2.7	1251	2	JH0256	botulinum neurotox
6	12	2.7	1252	2	S21178	botulinum neurotox
7	11	2.5	302	2	A55641	homeotic protein G
8	11	2.5	303	2	A56837	homeotic protein M
9	11	2.5	303	2	B49122	homeobox protein M
10	11	2.5	303	2	A48130	growth arrest-spec
11	11	2.5	354	2	S39406	homeotic protein o
12	11	2.5	355	2	S35345	otx1 protein - mou
13	11	2.5	355	2	I56547	homeodomain protei
14	11	2.5	367	2	S48106	neurotoxin type E
15	11	2.5	414	2	A40350	transcription repr
16	11	2.5	414	2	A48273	delta/YY1/NF-E1/UC
17	11	2.5	420	2	T39712	hypothetical prote
18	11	2.5	427	2	T42516	hypothetical prote
19	11	2.5	1180	2	S69205	stripe a/b protein
20	11	2.5	1296	1	BTCLAB	bontoxilysin (EC 3
21	11	2.5	1296	2	I40645	botulinum neurotox
22	10	2.2	132	2	I39193	gene HOXA1 protein
23	10	2.2	133	2	B30242	stem cell protein
24	10	2.2	222	2	I39192	gene HOXA1 protein
25	10	2.2	236	2	S41512	Brn-3b protein - m
26	10	2.2	298	2	S41469	homeotic protein M
27	10	2.2	309	2	T29293	hypothetical prote
28	10	2.2	321	2	T02987	myb-related protei
29	10	2.2	326	2	D83483	probable metal tra
30	10	2.2	331	2	A30242	homeotic protein E
31	10	2.2	335	2	G01448	homeobox protein H
32	10	2.2	351	1	KGZQHL	histidine-rich gly
33	10	2.2	375	2	A46390	cAMP receptor subt
34	10	2.2	387	2	A47446	HNF-3/fork head fa
35	10	2.2	390	1	A38565	polycomb (Pc) prot
36	10	2.2	391	2	H86187	hypothetical prote
37	10	2.2	407	2	G84783	probable pectinest
38	10	2.2	410	2	I38502	gene Brn-3b protei
39	10	2.2	411	2	I58156	Brn-3.2 - mouse
40	10	2.2	430	2	S66671	neuron-derived rec
41	10	2.2	451	2	A55909	transforming prote
42	10	2.2	469	2	I37451	HBF-G2 (HFK-2) pro
43	10	2.2	476	2	A54743	transcription fact
44	10	2.2	480	2	JH0672	brain factor 1 pro
45	10	2.2	580	2	T46024	hypothetical prote

ALIGNMENTS

RESULT 1
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific i
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 94.2%; Score 422; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
            ||
Db       1273 KE 1274


RESULT 2
S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin

  Query Match             33.5%;  Score 150;  DB 2;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-148;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 279

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 339

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNR 176
              ||||||||||||||||||||||||||||||
Db        340 TVRDCEIIWTLQDTSGNKENLIFRYEELNR 369


RESULT 3
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match              5.8%;  Score 26;  DB 2;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 1e-18;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         43 SILDMRYENNKFIDISGYGSNISING 68
              ||||||||||||||||||||||||||
Db        236 SILDMRYENNKFIDISGYGSNISING 261


RESULT 4
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match              2.9%;  Score 13;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 0.00012;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         71 YIYSTNRNQFGIY 83
              |||||||||||||
Db        889 YIYSTNRNQFGIY 901


RESULT 5
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match              2.7%;  Score 12;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        986 YINKWIFVTITN 997


RESULT 6
S21178
```

```
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match              2.7%;  Score 12;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        986 YINKWIFVTITN 997


RESULT 7
A55641
homeotic protein GAX - human
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 24-Sep-1999
C;Accession: A55641
R;LePage, D.F.; Altomare, D.A.; Testa, J.R.; Walsh, K.
Genomics 24, 535-540, 1994
A;Title: Molecular cloning and localization of the human GAX gene to 7p21.
A;Reference number: A55641; MUID:95229154; PMID:7713505
A;Accession: A55641
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <LEP>
A;Cross-references: GB:L36328; NID:g557549; PIDN:AAA58497.1; PID:g557550
C;Genetics:
A;Gene: GAX
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;186-242/Domain: homeobox homology <HOX>

  Query Match              2.5%;  Score 11;  DB 2;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 0.0038;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 8
A56837
homeotic protein MOX2 - human
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 24-Sep-1999
C;Accession: A56837; S52168
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.S.; Theodorakis, K.; Mankoo, B.; Pachnis, V.,
Genomics 26, 550-555, 1995
A;Title: Isolation of the human MOX2 homeobox gene and localization to chromosome 7p22.1
A;Reference number: A56837; MUID:95331791; PMID:7607679
A;Accession: A56837
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-303 <GRI>
A;Cross-references: EMBL:X82629; NID:g732790; PIDN:CAA57949.1; PID:g732791
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.; Theodorakis, K.; Maukoo, B.; Pachnis, V.; K
submitted to the EMBL Data Library, November 1994
A;Description: Isolation of the human MOX2 homeobox gene and localization to chromosome
A;Reference number: S52168
A;Accession: S52168
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-7,'R',9-110,'R',112-122,'Q',124-157,'V',159-303 <GR2>
A;Cross-references: EMBL:X82629
C;Genetics:
A;Gene: GDB:MOX2
A;Cross-references: GDB:120703; OMIM:155970
A;Map position: 3q13-3q13
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match              2.5%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0038;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 9
B49122
homeobox protein Mox-2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Dec-1993 #sequence_revision 17-Nov-1995 #text_change 17-Nov-2000
C;Accession: B49122; S41779; S29902
R;Candia, A.F.; Hu, J.; Crosby, J.; Lalley, P.A.; Noden, D.; Nadeau, J.H.; Wright, C.V.
Development 116, 1123-1136, 1992
```

```
A;Title: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially e
A;Reference number: A49122; MUID:93201999; PMID:1363541
A;Accession: B49122
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN1>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
A;Experimental source: C57BL
A;Note: the complete translation is not shown
R;Candia, A.F.; Kovalik, J.P.; Wright, C.V.E.
Nucleic Acids Res. 21, 4982, 1993
A;Title: Amino acid sequence of Mox-2 and comparison to its Xenopus and rat homologs.
A;Reference number: S41469; MUID:94232829; PMID:7909944
A;Accession: S41779
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN2>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
R;Candia, A.
submitted to the EMBL Data Library, October 1992
A;Reference number: S29902
A;Accession: S29902
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 186-246 <CAN>
A;Cross-references: EMBL:Z16406
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.5%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0038;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 10
A48130
growth arrest-specific homeobox protein Cvx - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A48130; S31976
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
Mol. Cell. Biol. 13, 3722-3733, 1993
A;Title: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated du
A;Reference number: A48130; MUID:93268321; PMID:8098844
A;Accession: A48130
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-303 <GOR>
A;Cross-references: GB:Z17223; NID:g57951; PIDN:CAA78931.1; PID:g57952
A;Experimental source: aorta
A;Note: sequence extracted from NCBI backbone (NCBIN:132842, NCBIP:132843)
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
submitted to the EMBL Data Library, October 1992
A;Description: Cvx: A diverged homeodomain gene which is rapidly down-regulated followin
A;Reference number: S31976
A;Accession: S31976
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-182,'S',184-303 <GO2>
A;Cross-references: EMBL:Z17223
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

  Query Match             2.5%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0038;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 11
S39406
homeotic protein otx1 - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 15-Oct-1999
C;Accession: S39406
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; B
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S39406
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-354 <SIM>
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

  Query Match             2.5%;  Score 11;  DB 2;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.0044;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 12
S35345
otx1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000
C;Accession: S35345; S29087
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; B
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S35345
A;Molecule type: DNA
A;Residues: 1-355 <SIM>
R;Simeone, A.; Acampora, D.; Gulisano, M.; Stornaiuolo, A.; Boncinelli, E.
Nature 358, 687-690, 1992
A;Title: Nested expression domains of four homeobox genes in developing rostral brain.
A;Reference number: S29084; MUID:92357147; PMID:1353865
A;Accession: S29087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 31-98 <SI2>
A;Cross-references: EMBL:X68883; NID:g53540; PIDN:CAA48754.1; PID:g53541
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

  Query Match             2.5%;  Score 11;  DB 2;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0044;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 13
I56547
```

homeodomain protein otx1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C;Accession: I56547
R;Frantz, G.D.; Weimann, J.M.; Levin, M.E.; McConnell, S.K.
J. Neurosci. 14, 5725-5740, 1994
A;Title: Otx1 and Otx2 define layers and regions in developing cerebral cortex and cereb
A;Reference number: I56547; MUID:95016961; PMID:7931541
A;Accession: I56547
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: GB:L32602; NID:g535739; PIDN:AAA53557.1; PID:g535740
C;Genetics:
A;Gene: OTX1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.5%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 14
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 2.5%; Score 11; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 106 NFSISFWVRIP 116
||||||||||
Db 299 NFSISFWVRIP 309

RESULT 15
A40350
transcription repressor protein YY1 - human
N;Alternate names: transcription repressor/activator protein NF-E1
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Oct-1999
C;Accession: A40350; S78494; S33712; A56419
R;Shi, Y.; Seto, E.; Chang, L.S.; Shenk, T.
Cell 67, 377-388, 1991
A;Title: Transcriptional repression by YY1, a human GLI-Krueppel-related protein, and re
A;Reference number: A40350; MUID:92005716; PMID:1655281
A;Accession: A40350
A;Molecule type: mRNA
A;Residues: 1-414 <SHI>
A;Cross-references: GB:M77698; NID:g186767; PIDN:AAA59467.1; PID:g186768
A;Experimental source: HeLa cells
A;Note: the authors translated the codon CGC for residue 371 as Lys, CGA for residue 375

R;Whitson, R.H.; Huang, T.; Dang, J.; Itakura, K.
submitted to the EMBL Data Library, July 1992
A;Description: Observed and predicted DNA binding of a zinc finger protein which recogn
A;Reference number: S78494
A;Accession: S78494
A;Molecule type: mRNA
A;Residues: 1-195,'G',197-414 <WHI>
A;Cross-references: EMBL:Z14077; NID:g38010; PIDN:CAA78455.1; PID:g38011
R;Park, K.; Atchison, M.L.
Proc. Natl. Acad. Sci. U.S.A. 88, 9804-9808, 1991
A;Title: Isolation of a candidate repressor/activator, NF-E1 (YY-1), delta, that binds t
A;Reference number: A56419; MUID:92052179; PMID:1946405
A;Accession: S33712
A;Molecule type: mRNA
A;Residues: 1-64,'R',66-195,'G',197-414 <PAR>
A;Cross-references: GB:M76541; NID:g189173; PIDN:AAA59926.1; PID:g189174
C;Keywords: DNA binding; transcription regulation; zinc finger
F;298-320/Region: zinc finger CCHH motif
F;327-347/Region: zinc finger CCHH motif
F;355-377/Region: zinc finger CCHH motif
F;385-407/Region: zinc finger CCHH motif

Query Match 2.5%; Score 11; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 69 GHHHHHHHHHH 79

Search completed: September 10, 2004, 05:14:54
Job time : 7.30641 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 3.87189 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 448
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	422	94.2	1274	1	BXF_CLOBO	P30996 clostridium
2	12	2.7	1250	1	BXE_CLOBO	Q00496 clostridium
3	12	2.7	1250	1	BXE_CLOBU	P30995 clostridium
4	11	2.5	303	1	MOX2_HUMAN	P50222 homo sapien
5	11	2.5	303	1	MOX2_MOUSE	P32443 mus musculu
6	11	2.5	303	1	MOX2_RAT	P39020 rattus norv
7	11	2.5	354	1	OTX1_HUMAN	P32242 homo sapien
8	11	2.5	355	1	OTX1_MOUSE	P80205 mus musculu
9	11	2.5	355	1	OTX1_RAT	Q63410 rattus norv
10	11	2.5	414	1	TYY1_HUMAN	P25490 homo sapien
11	11	2.5	414	1	TYY1_MOUSE	Q00899 mus musculu
12	11	2.5	420	1	YBE1_SCHPO	O42980 schizosacch
13	11	2.5	466	1	ZIC3_MOUSE	Q62521 mus musculu
14	11	2.5	467	1	ZIC3_HUMAN	O60481 homo sapien
15	11	2.5	611	1	SNF1_CANGA	Q00372 candida gla
16	11	2.5	619	1	SNF1_CANTR	O94168 candida tro
17	11	2.5	1295	1	BXA1_CLOBO	P10845 clostridium
18	11	2.5	1295	1	BXA2_CLOBO	Q45894 clostridium
19	10	2.2	202	1	HAN1_CHICK	Q90691 gallus gall
20	10	2.2	204	1	HAN1_SHEEP	Q28555 ovis aries
21	10	2.2	298	1	MOX2_XENLA	P39021 xenopus lae
22	10	2.2	331	1	HXA1_MOUSE	P09022 mus musculu
23	10	2.2	335	1	HXA1_HUMAN	P49639 homo sapien
24	10	2.2	351	1	HRPX_PLALO	P04929 plasmodium
25	10	2.2	375	1	CAR2_DICDI	P34907 dictyosteli
26	10	2.2	387	1	QIN_AVIS3	P56260 avian sarco
27	10	2.2	390	1	PC_DROME	P26017 drosophila
28	10	2.2	410	1	PO42_HUMAN	Q12837 homo sapien
29	10	2.2	411	1	PO42_MOUSE	Q63934 mus musculu
30	10	2.2	430	1	NOR2_RAT	Q63516 rattus norv
31	10	2.2	449	1	GAT6_HUMAN	Q92908 homo sapien
32	10	2.2	451	1	FXGB_CHICK	Q90964 gallus gall
33	10	2.2	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
34	10	2.2	465	1	HNF6_MOUSE	O08755 mus musculu
35	10	2.2	465	1	HNF6_RAT	P70512 rattus norv
36	10	2.2	469	1	FXGA_HUMAN	P55316 homo sapien
37	10	2.2	477	1	FXGB_HUMAN	P55315 homo sapien
38	10	2.2	480	1	FXGB_RAT	Q00939 rattus norv
39	10	2.2	481	1	FXGB_MOUSE	Q60987 mus musculu
40	10	2.2	551	1	CBX4_MOUSE	O55187 mus musculu
41	10	2.2	558	1	CBX4_HUMAN	O00257 homo sapien
42	10	2.2	620	1	SNF1_CANAL	P52497 candida alb
43	10	2.2	623	1	PNT1_DROME	P51022 drosophila
44	10	2.2	626	1	NR43_HUMAN	Q92570 homo sapien
45	10	2.2	628	1	NR43_RAT	P51179 rattus norv

ALIGNMENTS

RESULT 1
BXF_CLOBO
ID BXF_CLOBO STANDARD; PRT; 1274 AA.
AC P30996;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE (Bontoxilysin F).
GN BOTF.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23387;
RX MEDLINE=93012902; PubMed=1398040;
RA East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA Roberts T.A., Thompson D.E.;
RT "Sequence of the gene encoding type F neurotoxin of Clostridium
RT botulinum.";
RL FEMS Microbiol. Lett. 75:225-230(1992).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=Hobbs FT10;
RX MEDLINE=94297488; PubMed=7764998;
RA East A.K., Collins M.D.;
RT "Conserved structure of genes encoding components of botulinum
RT neurotoxin complex M and the sequence of the gene coding for the
RT nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL Curr. Microbiol. 29:69-77(1994).
RN [3]
RP SEQUENCE OF 634-1002 FROM N.A.
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinal neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC BOND OF SYNAPTOBREVINS-1 AND -2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No

```
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             94.2%;  Score 422;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
              ||
Db       1273 KE 1274


RESULT 2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
```

```
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             2.7%;  Score 12;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0.0061;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        985 YINKWIFVTITN 996


RESULT 3
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
```

```
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             2.7%;  Score 12;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0.0061;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        985 YINKWIFVTITN 996


RESULT 4
MOX2_HUMAN
ID   MOX2_HUMAN     STANDARD;      PRT;   303 AA.
AC   P50222; Q9UPL6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE   homeobox).
GN   MEOX2 OR MOX2 OR GAX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95331791; PubMed=7607679;
RA   Grigoriou M., Kastrinaki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA   Pachnis V., Karagogeos D.;
RT   "Isolation of the human MOX2 homeobox gene and localization to
RT   chromosome 7p22.1-p21.3.";
RL   Genomics 26:550-555(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95229154; PubMed=7713505;
RA   Lepage D.F., Walsh K.;
RT   "Molecular cloning and localization of the human GAX gene to 7p21.";
RL   Genomics 24:535-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 230-303 FROM N.A.
RA   Cordes M., Lacy M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Embryo and placenta.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X82629; CAA57949.1; -.
DR   EMBL; L36328; AAA58497.1; -.
DR   EMBL; BC017021; AAH17021.1; -.
DR   EMBL; AC004452; AAC06184.1; -.
DR   PIR; A55641; A55641.
DR   PIR; A56837; A56837.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04005; -.
DR   Genew; HGNC:7014; MEOX2.
DR   MIM; 600535; -.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DNA_BIND    186    245       HOMEOBOX.
FT   CONFLICT     58     58       G -> D (IN REF. 2).
FT   CONFLICT     79     79       MISSING (IN REF. 2).
SQ   SEQUENCE   303 AA;  33457 MW;  809ADE0CD090023D CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.015;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
MOX2_MOUSE
ID   MOX2_MOUSE     STANDARD;      PRT;   303 AA.
AC   P32443;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201999; PubMed=1363541;
RA   Candia A.F., Hu J., Crosby J., Lalley P.A., Noden D., Nadeau J.H.,
RA   Wright C.V.E.;
RT   "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are
RT   differentially expressed during early mesodermal patterning in mouse
RT   embryos.";
RL   Development 116:1123-1136(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94232829; PubMed=7909944;
RA   Candia A.F., Kovalik J.-P., Wright C.V.E.;
RT   "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat
RT   homologs.";
RL   Nucleic Acids Res. 21:4982-4982(1993).
RN   [3]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=95349593; PubMed=7623821;
RA   Andres V., Fisher S., Wearsch P., Walsh K.;
RT   "Regulation of Gax homeobox gene transcription by a combination of
RT   positive factors including myocyte-specific enhancer factor 2.";
RL   Mol. Cell. Biol. 15:4272-4281(1995).
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-
CC       8.5 dpc it is found on the entire epithelium of the somite. At 9.5
CC       dpc its expression is restricted to the sclerotome. At 10.5 dpc it
CC       is found in sclerotomally derived cells including the vertebral
CC       and costal precursors.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16406; CAA78899.1; -.
DR   EMBL; S79168; -; NOT_ANNOTATED_CDS.
DR   PIR; B49122; B49122.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04048; -.
DR   MGD; MGI:103219; Meox2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       63     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33506 MW;  41BD05FC39AA4427 CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.015;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 6
MOX2_RAT
ID   MOX2_RAT       STANDARD;      PRT;   303 AA.
AC   P39020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Growth arrest-specific homeobox).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93268321; PubMed=8098844;
RA   Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,
RA   Walsh K.;
RT   "Molecular cloning of a diverged homeobox gene that is rapidly down-
RT   regulated during the G0/G1 transition in vascular smooth muscle
RT   cells.";
RL   Mol. Cell. Biol. 13:3722-3733(1993).
RN   [2]
RP   REVISIONS.
RA   Walsh K.;
RL   Submitted (MAR-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Aorta and heart. Also detected in lung and
CC       kidney.
CC   -!- INDUCTION: Rapidly and transiently down-regulated during the
CC       transition from G0 to G1 induced by mitogen stimulation.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z17223; CAA78931.1; -.
DR   PIR; A48130; A48130.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04054; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       64     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33605 MW;  7776642AEFA3A2E8 CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 303;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0.015;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 7
OTX1_HUMAN
ID   OTX1_HUMAN     STANDARD;      PRT;   354 AA.
AC   P32242;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS):
CC       5'-TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC007621; AAH07621.1; -.
DR   PIR; S39406; S39406.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02079; -.
DR   Genew; HGNC:8521; OTX1.
DR   MIM; 600036; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    301       HIS-RICH.
SQ   SEQUENCE   354 AA;  37327 MW;  E32C1E47464BDDAA CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
OTX1_MOUSE
ID   OTX1_MOUSE     STANDARD;      PRT;   355 AA.
AC   P80205;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1 OR OTX-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
```

```
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 31-98 FROM N.A.
RX   MEDLINE=92357147; PubMed=1353865;
RA   Simeone A., Acampora D., Gulisano M., Stornaiuolo A., Boncinelli E.;
RT   "Nested expression domains of four homeobox genes in developing
RT   rostral brain.";
RL   Nature 358:687-690(1992).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC057105; AAH57105.1; -.
DR   EMBL; BC058354; AAH58354.1; -.
DR   EMBL; X68883; CAA48754.1; -.
DR   PIR; S35345; S35345.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02080; -.
DR   MGD; MGI:97450; Otx1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
SQ   SEQUENCE   355 AA;  37531 MW;  DA5136D9D4FC948F CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 9
OTX1_RAT
ID   OTX1_RAT       STANDARD;      PRT;   355 AA.
AC   Q63410; Q64203;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95016961; PubMed=7931541;
RA   Frantz G.D., Weimann J.M., Levin M.E., McConnell S.K.;
RT   "Otx1 and Otx2 define layers and regions in developing cerebral
RT   cortex and cerebellum.";
RL   J. Neurosci. 14:5725-5740(1994).
RN   [2]
RP   SEQUENCE OF 176-323 FROM N.A.
RX   MEDLINE=96108898; PubMed=8613727;
RA   Robel L., Ding M., James A.J., Lin X., Simeone A., Leckman J.F.,
RA   Vaccarino F.M.;
RT   "Fibroblast growth factor 2 increases Otx2 expression in precursor
RT   cells from mammalian telencephalon.";
RL   J. Neurosci. 15:7879-7891(1995).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3' (By similarity). May play a role in the specification
CC       or differentiation of neurons in the deep layers of the cerebral
CC       cortex, and also in cerebellar regionalization during early
CC       development.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: Expressed throughout the forebrain and
CC       midbrain during development, and in addition is also seen in
CC       discrete spatial and temporal domains in the developing cerebral
CC       cortex and cerebellum. Confined to a subpopulation of neurons in
CC       layers 5 and 6 within the adult cerebral cortex and during
CC       development expression is high in the progenitors of these deep-
CC       layer cells. Expressed in the developing cerebellum in spatially
CC       restricted regions of the external granular layer.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32602; AAA53557.1; -.
DR   EMBL; S81924; -; NOT_ANNOTATED_CDS.
DR   PIR; I56547; I56547.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
FT   CONFLICT    219    219       T -> A (IN REF. 2).
FT   CONFLICT    229    229       R -> G (IN REF. 2).
```

```
SQ   SEQUENCE   355 AA;  37602 MW;  C875871723D0B876 CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 10
TYY1_HUMAN
ID   TYY1_HUMAN     STANDARD;      PRT;   414 AA.
AC   P25490; Q14935;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1).
GN   YY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92005716; PubMed=1655281;
RA   Shi Y., Seto E., Chang L.-S., Shenk T.;
RT   "Transcriptional repression by YY1, a human GLI-Kruppel-related
RT   protein, and relief of repression by adenovirus E1A protein.";
RL   Cell 67:377-388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=92052179; PubMed=1946405;
RA   Park K., Atchison M.;
RT   "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
RT   that binds to the immunoglobulin kappa 3' enhancer and the
RT   immunoglobulin heavy-chain mu E1 site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Whitson R.H., Huang T., Dang J., Itakura K.;
RL   Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98152931; PubMed=9493912;
RA   McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
RA   Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
RT   "Targeting of the YY1 transcription factor to the nucleolus and the
RT   nuclear matrix in situ: the C-terminus is a principal determinant for
RT   nuclear trafficking.";
RL   J. Cell. Biochem. 68:500-510(1998).
RN   [6]
RP   INTERACTION WITH YAF2.
RX   MEDLINE=97169297; PubMed=9016636;
RA   Kalenik J.L., Chen D., Bradley M.E., Chen S.-J., Lee T.-C.;
RT   "Yeast two-hybrid cloning of a novel zinc finger protein that
RT   interacts with the multifunctional transcription factor YY1.";
RL   Nucleic Acids Res. 25:843-849(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
RX   MEDLINE=97098436; PubMed=8942976;
RA   Houbaviy H.B., Usheva A., Shenk T., Burley S.K.;
RT   "Cocrystal structure of YY1 bound to the adeno-associated virus P5
RT   initiator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 353-379.
RX   MEDLINE=98308000; PubMed=9642075;
RA   Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
RA   Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
RT   "Design, synthesis and structure of a zinc finger with an artificial
RT   beta-turn.";
RL   J. Mol. Biol. 279:973-986(1998).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
CC       EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
CC       PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77698; AAA59467.1; -.
DR   EMBL; M76541; AAA59926.1; -.
DR   EMBL; Z14077; CAA78455.1; -.
DR   EMBL; BC037308; AAH37308.1; -.
DR   PIR; A40350; A40350.
DR   PDB; 1UBD; 23-DEC-96.
DR   PDB; 1ZNM; 01-APR-98.
DR   TRANSFAC; T00915; -.
DR   Genew; HGNC:12856; YY1.
DR   MIM; 600013; -.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
```

```
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     69       GLY-RICH.
FT   DOMAIN       70     80       POLY-HIS.
FT   DOMAIN      159    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE 1.
FT   ZN_FING     325    347       C2H2-TYPE 2.
FT   ZN_FING     353    377       C2H2-TYPE 3.
FT   ZN_FING     383    407       C2H2-TYPE 4.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION.
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN.
FT   CONFLICT     65     65       H -> R (IN REF. 2).
FT   CONFLICT    196    196       G -> R (IN REF. 1).
FT   STRAND      296    297
FT   TURN        301    302
FT   STRAND      306    307
FT   HELIX       310    317
FT   HELIX       318    320
FT   STRAND      325    326
FT   TURN        328    330
FT   STRAND      333    334
FT   HELIX       337    343
FT   HELIX       344    346
FT   TURN        347    347
FT   STRAND      353    354
FT   TURN        358    359
FT   STRAND      363    364
FT   HELIX       367    378
FT   TURN        388    389
FT   TURN        397    400
FT   HELIX       401    407
SQ   SEQUENCE   414 AA;  44712 MW;  058C05A0AD2D04E6 CRC64;

  Query Match          2.5%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity  100.0%;  Pred. No. 0.019;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 11
TYY1_MOUSE
ID   TYY1_MOUSE     STANDARD;      PRT;   414 AA.
AC   Q00899;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1) (UCR-motif DNA-binding protein).
GN   YY1 OR UCRBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107191; PubMed=1309593;
RA   Flanagan J.R., Becker K.G., Ennist D.L., Gleason S.L., Driggers P.H.,
RA   Levi B.-Z., Appella E., Ozato K.;
RT   "Cloning of a negative transcription factor that binds to the
RT   upstream conserved region of Moloney murine leukemia virus.";
RL   Mol. Cell. Biol. 12:38-44(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93296177; PubMed=8516301;
RA   Safrany G., Perry R.P.;
RT   "Characterization of the mouse gene that encodes the delta/YY1/NF-
RT   E1/UCRBP transcription factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5559-5563(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052178; PubMed=1946404;
RA   Hariharan N., Kelley D.E., Perry R.P.;
RT   "Delta, a transcription factor that binds to downstream elements in
RT   several polymerase II promoters, is a functionally versatile zinc
RT   finger protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9799-9803(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. BINDS TO
CC       THE UPSTREAM CONSERVED REGION (UCR) (5'-CGCCATTTT-3') OF MOLONEY
CC       MURINE LEUKEMIA VIRUS (MULV).
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73963; AAA40522.1; -.
DR   EMBL; L13968; AAA40477.1; -.
DR   EMBL; L13969; AAA40477.1; JOINED.
DR   EMBL; L13965; AAA40477.1; JOINED.
DR   EMBL; L13966; AAA40477.1; JOINED.
DR   EMBL; L13967; AAA40477.1; JOINED.
DR   EMBL; M74590; AAA37521.1; -.
DR   EMBL; BC055899; AAH55899.1; -.
DR   PIR; A48273; A48273.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T00278; -.
DR   TRANSFAC; T00865; -.
DR   MGD; MGI:99150; Yy1.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003700; F:transcription factor activity; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR007087; Znf_C2H2.
```

```
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     70       GLY-RICH.
FT   DOMAIN       71     82       POLY-HIS.
FT   DOMAIN      161    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    407       C2H2-TYPE.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION
FT                                (BY SIMILARITY).
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION (BY SIMILARITY).
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN (BY SIMILARITY).
FT   CONFLICT    219    219       F -> S (IN REF. 3).
FT   CONFLICT    375    375       R -> G (IN REF. 3).
SQ   SEQUENCE   414 AA;  44717 MW;  C012378288E984F9 CRC64;

  Query Match              2.5%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.019;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 12
YBE1_SCHPO
ID   YBE1_SCHPO     STANDARD;      PRT;   420 AA.
AC   O42980; P78807;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative thiol protease C17D11.01 (EC 3.4.22.-).
GN   SPBC17D11.01 OR SPBC20F10.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- SIMILARITY: Belongs to peptidase family C48.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031739; CAA21072.1; -.
DR   EMBL; AL021747; CAA16851.1; -.
DR   EMBL; D89156; BAA13818.1; -.
DR   GeneDB_SPombe; SPBC17D11.01; -.
DR   InterPro; IPR003653; Peptidase_C48.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Thiol protease.
FT   DOMAIN      299    312       POLY-HIS.
FT   CONFLICT     57     57       L -> W (IN REF. 2).
FT   CONFLICT    420    420       F -> FGCYTCVV (IN REF. 2).
SQ   SEQUENCE   420 AA;  47257 MW;  2125D94D9A2D9E87 CRC64;

  Query Match              2.5%;  Score 11;  DB 1;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 13
ZIC3_MOUSE
ID   ZIC3_MOUSE     STANDARD;      PRT;   466 AA.
AC   Q62521;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=96132843; PubMed=8557628;
RA   Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA   Chapman V.M., Mikoshiba K.;
RT   "The mouse zic gene family. Homologues of the Drosophila pair-rule
RT   gene odd-paired.";
RL   J. Biol. Chem. 271:1043-1047(1996).
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: CNS. A high level expression is seen in the
CC       cerebellum.
```

```
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D70849; BAA11116.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T04671; -.
DR   MGD; MGI:106676; Zic3.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     96       POLY-HIS.
FT   ZN_FING     294    321       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     327    351       C2H2-TYPE 2.
FT   ZN_FING     357    381       C2H2-TYPE 3.
FT   ZN_FING     387    409       C2H2-TYPE 4.
SQ   SEQUENCE   466 AA;  50644 MW;  089749DF726F069B CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 0.021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 14
ZIC3_HUMAN
ID   ZIC3_HUMAN     STANDARD;      PRT;   467 AA.
AC   O60481;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HTX1 MET-323.
RX   MEDLINE=98016414; PubMed=9354794;
RA   Gebbia M., Ferrero G.B., Pilia G., Bassi M.T., Aylsworth A.S.,
RA   Penman-Splitt M., Bird L.M., Bamforth J.S., Burn J., Schlessiner D.,
RA   Nelson D.L., Casey B.;
RT   "X-linked situs abnormalities result from mutations in ZIC3.";
RL   Nat. Genet. 17:305-308(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in ZIC3 are the cause of X-Linked visceral
CC       heterotaxy (HTX1) [MIM:306955]. HTX1 is a disease characterized by
CC       congenital heart disease and alterations of visceral situs,
CC       including asplenia or polysplenia, symmetric liver, intestinal
CC       malrotation, and abnormal lung lobation.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028706; AAC05594.1; -.
DR   EMBL; AL035443; CAB41648.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:12874; ZIC3.
DR   MIM; 300265; -.
DR   MIM; 306955; -.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein; Disease mutation.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     97       POLY-HIS.
FT   ZN_FING     295    322       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     328    352       C2H2-TYPE 2.
FT   ZN_FING     358    382       C2H2-TYPE 3.
FT   ZN_FING     388    410       C2H2-TYPE 4.
FT   VARIANT     323    323       T -> M (in HTX1).
FT                                /FTId=VAR_007753.
SQ   SEQUENCE   467 AA;  50569 MW;  3150CF13C0679568 CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 15
SNF1_CANGA
ID   SNF1_CANGA     STANDARD;      PRT;   611 AA.
AC   Q00372;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCCLS84;
RX   MEDLINE=97101049; PubMed=8945576;
RA   Petter R., Kwon-Chung K.J.;
RT   "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT   pathogenic yeast Candida glabrata.";
RL   Infect. Immun. 64:5269-5273(1996).
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
```

```
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78130; AAB48642.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN        6     17       POLY-HIS.
FT   DOMAIN       39    290       PROTEIN KINASE.
FT   NP_BIND      45     53       ATP (BY SIMILARITY).
FT   BINDING      68     68       ATP (BY SIMILARITY).
FT   ACT_SITE    161    161       BY SIMILARITY.
FT   MOD_RES     194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  70049 MW;  89E17812A4900CD0 CRC64;

  Query Match             2.5%;  Score 11;  DB 1;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 0.027;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db          8 HHHHHHHHHHS 18


Search completed: September 10, 2004, 05:07:22
Job time : 4.87189 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 18.2206 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 448
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	19.6	1280	2	Q9ZAJ5	Q9zaj5 clostridium
2	27	6.0	1278	2	Q57236	Q57236 clostridium
3	13	2.9	1268	2	Q45851	Q45851 clostridium
4	12	2.7	226	11	Q9QZD2	Q9qzd2 rattus norv
5	12	2.7	516	13	O42186	O42186 brachydanio
6	12	2.7	517	13	Q90YL6	Q90yl6 brachydanio
7	12	2.7	569	13	Q802D2	Q802d2 brachydanio
8	12	2.7	679	13	Q803K9	Q803k9 brachydanio
9	12	2.7	706	5	Q8SXG1	Q8sxg1 drosophila
10	12	2.7	861	5	Q9VRT8	Q9vrt8 drosophila
11	12	2.7	1251	2	Q9K395	Q9k395 clostridium
12	12	2.7	1252	2	Q8KZM3	Q8kzm3 clostridium
13	12	2.7	1255	2	Q9FAR6	Q9far6 clostridium
14	11	2.5	163	5	Q7YTA3	Q7yta3 glomeris ma
15	11	2.5	225	13	Q90XZ3	Q90xz3 erpetoichth
16	11	2.5	230	4	O75263	O75263 homo sapien
17	11	2.5	243	4	Q9P2R2	Q9p2r2 homo sapien
18	11	2.5	243	6	Q9N2G2	Q9n2g2 gorilla gor
19	11	2.5	243	6	Q9N2G3	Q9n2g3 pan troglod
20	11	2.5	243	6	Q9N2G1	Q9n2g1 pongo pygma
21	11	2.5	302	13	Q90YH7	Q90yh7 gallus gall
22	11	2.5	303	6	Q95JA6	Q95ja6 sus scrofa
23	11	2.5	303	11	Q99M23	Q99m23 mus musculu
24	11	2.5	355	11	Q91ZJ3	Q91zj3 mus musculu
25	11	2.5	358	13	Q9DG32	Q9dg32 xenopus lae
26	11	2.5	367	2	Q45861	Q45861 clostridium
27	11	2.5	367	2	Q45862	Q45862 clostridium
28	11	2.5	411	11	Q8CHJ4	Q8chj4 rattus norv
29	11	2.5	441	5	Q8T1W3	Q8t1w3 dictyosteli
30	11	2.5	449	10	Q8L4M9	Q8l4m9 oryza sativ
31	11	2.5	456	11	Q8BSB3	Q8bsb3 mus musculu
32	11	2.5	467	13	Q9YGC6	Q9ygc6 xenopus lae
33	11	2.5	511	10	Q8GXZ5	Q8gxz5 arabidopsis
34	11	2.5	906	5	Q8IN91	Q8in91 drosophila
35	11	2.5	1122	5	Q8SXD4	Q8sxd4 drosophila
36	11	2.5	1147	5	Q95PH8	Q95ph8 dictyosteli
37	11	2.5	1180	5	Q24163	Q24163 drosophila
38	11	2.5	1186	5	Q9VED3	Q9ved3 drosophila
39	11	2.5	1283	5	Q95RH4	Q95rh4 drosophila
40	11	2.5	1911	5	Q9W4M7	Q9w4m7 drosophila
41	10	2.2	53	2	O55244	O55244 escherichia
42	10	2.2	233	4	Q8TC53	Q8tc53 homo sapien
43	10	2.2	233	5	Q8MP30	Q8mp30 dictyosteli
44	10	2.2	236	11	Q9QWE1	Q9qwe1 mus sp. cla
45	10	2.2	239	5	Q7Z1G6	Q7z1g6 trypanosoma

ALIGNMENTS

RESULT 1
Q9ZAJ5
ID Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.

```
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             19.6%;  Score 88;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-82;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIR 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 YINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIR 1066

Qy        240 YFKVFNTELDKTEIETLYSNEPDPSILK 267
              ||||||||||||||||||||||||||||
Db       1067 YFKVFNTELDKTEIETLYSNEPDPSILK 1094


RESULT 2
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match              6.0%;  Score 27;  DB 2;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-18;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||
Db       1004 YINKWIFVTITNNRLGNSRIYINGNLI 1030


RESULT 3
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match              2.9%;  Score 13;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 0.00044;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         71 YIYSTNRNQFGIY 83
              |||||||||||||
Db        889 YIYSTNRNQFGIY 901


RESULT 4
Q9QZD2
ID   Q9QZD2      PRELIMINARY;      PRT;   226 AA.
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
SQ   SEQUENCE   226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

  Query Match             2.7%;  Score 12;  DB 11;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 0.0011;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db         17 HHHHHHHHHHSS 28


RESULT 5
O42186
ID   O42186      PRELIMINARY;      PRT;   516 AA.
AC   O42186;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FTZ-F1 homolog.
GN   NR5A2 OR ZFF1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322094; PubMed=9178748;
RA   Liu D., Le Drean Y., Ekker M., Xiong F., Hew C.L.;
RT   "Teleost FTZ-F1 homolog and its splicing variant determine the
RT   expression of the salmon gonadotropin IIbeta subunit gene.";
RL   Mol. Endocrinol. 11:877-890(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AF014926; AAC60274.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04065; -.
DR   ZFIN; ZDB-GENE-990415-79; nr5a2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   516 AA;  58805 MW;  500DA09E116D5FA5 CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 6
Q90YL6
ID   Q90YL6      PRELIMINARY;      PRT;   517 AA.
AC   Q90YL6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nuclear receptor FF1aE.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chai C., Tan J.H., Gao W., Teh H.L., Chan W.K.;
RT   "Full length zebrafish ff1aE cDNA.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AY034061; AAK54449.1; -.
DR   HSSP; P19793; 1LBD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   517 AA;  58761 MW;  CF48D08AB7A75F29 CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 7
Q802D2
ID   Q802D2      PRELIMINARY;      PRT;   569 AA.
AC   Q802D2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
```

```
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048886; AAH48886.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
KW   Cyclin.
SQ   SEQUENCE   569 AA;  64117 MW;  812A7332C06EF4C2 CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 0.0024;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        486 HHHHHHHHHHSS 497


RESULT 8
Q803K9
ID   Q803K9      PRELIMINARY;      PRT;   679 AA.
AC   Q803K9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T1 (Fragment).
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044435; AAH44435.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
KW   Cyclin.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  74887 MW;  AB5667D59439659F CRC64;

  Query Match             2.7%;  Score 12;  DB 13;  Length 679;
  Best Local Similarity   100.0%;  Pred. No. 0.0028;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        488 HHHHHHHHHHSS 499


RESULT 9
Q8SXG1
ID   Q8SXG1      PRELIMINARY;      PRT;   706 AA.
AC   Q8SXG1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RH12258p.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AY089654; AAL90392.1; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   706 AA;  79683 MW;  73C0F8199DF60E5F CRC64;

  Query Match             2.7%;  Score 12;  DB 5;  Length 706;
  Best Local Similarity   100.0%;  Pred. No. 0.0029;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        370 GHHHHHHHHHHS 381


RESULT 10
Q9VRT8
ID   Q9VRT8      PRELIMINARY;      PRT;   861 AA.
AC   Q9VRT8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG6619 protein.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
```

```
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AE003564; AAF50702.2; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   861 AA;  96809 MW;  8530ED5D92671A41 CRC64;

  Query Match             2.7%;  Score 12;  DB 5;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 0.0034;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        525 GHHHHHHHHHHS 536


RESULT 11
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
```

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             2.7%;  Score 12;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0.0048;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        986 YINKWIFVTITN 997


RESULT 12
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match             2.7%;  Score 12;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0.0048;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        986 YINKWIFVTITN 997


RESULT 13
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match             2.7%;  Score 12;  DB 2;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 0.0048;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        180 YINKWIFVTITN 191
              ||||||||||||
Db        989 YINKWIFVTITN 1000


RESULT 14
Q7YTA3
ID   Q7YTA3      PRELIMINARY;      PRT;   163 AA.
AC   Q7YTA3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Distal-less protein (Fragment).
GN   DLL.
OS   Glomeris marginata.
OC   Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Pentazonia;
OC   Glomerida; Glomeridae; Glomeris.
OX   NCBI_TaxID=62006;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22767835; PubMed=12885558;
RA   Prpic N.M., Tautz D.;
RT   "The expression of the proximodistal axis patterning genes Distal-less
RT   and dach shund in the appendages of Glomeris marginata (Myriapoda:
RT   Diplopoda) suggests a special role for these genes in patterning the
RT   head appendages.";
RL   Dev. Biol. 260:97-112(2003).
DR   EMBL; AJ551276; CAD82905.1; -.
FT   NON_TER       1      1
FT   NON_TER     163    163
SQ   SEQUENCE   163 AA;  17588 MW;  87B8BCD9F5421697 CRC64;

  Query Match             2.5%;  Score 11;  DB 5;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHH 11
              ||||||||||||
Db         16 MGHHHHHHHHH 26


RESULT 15
Q90XZ3
```

```
ID   Q90XZ3      PRELIMINARY;      PRT;   225 AA.
AC   Q90XZ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Homeodomain protein Otx1 (Fragment)..
OS   Erpetoichthys calabaricus (Rope fish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Polypteriformes; Polypteridae; Erpetoichthys.
OX   NCBI_TaxID=27687;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21396394; PubMed=11504847;
RA   Germot A., Lecointre G., Plouhinec J.L., Le Mentec C., Girardot F.,
RA   Mazan S.;
RT   "Structural evolution of otx genes in Craniates.";
RL   Mol. Biol. Evol. 18:1668-1678(2001).
DR   EMBL; AF321406; AAL01673.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     225    225
SQ   SEQUENCE   225 AA;  23475 MW;  0AE78B8DB48B4941 CRC64;

  Query Match             2.5%;  Score 11;  DB 13;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        172 GHHHHHHHHHH 182


Search completed: September 10, 2004, 05:12:57
Job time : 19.2206 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 29.5516 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 2404
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2404	100.0	448	2	AAW68399	Aaw68399 Clostridi
2	1800.5	74.9	862	4	AAE07890	Aae07890 Modified
3	1800.5	74.9	887	4	AAE07892	Aae07892 Modified
4	1800.5	74.9	979	6	AAE35713	Aae35713 BoNT/F-Hc
5	1800.5	74.9	1192	6	AAE35711	Aae35711 BoNT/F-Hc
6	1800.5	74.9	1192	6	AAE35710	Aae35710 BoNT/F-Hc
7	1800	74.9	431	2	AAW09014	Aaw09014 Immunogen
8	1800	74.9	432	4	AAB04103	Aab04103 Botulism
9	1800	74.9	432	4	AAB04096	Aab04096 Botulism
10	1800	74.9	645	4	AAE07894	Aae07894 Modified
11	1800	74.9	645	6	AAE35692	Aae35692 DipT HN d
12	1800	74.9	657	6	AAE35693	Aae35693 BoNT/F-Hc
13	1800	74.9	657	6	AAE35694	Aae35694 BoNT/F-Hc
14	1800	74.9	685	4	AAE07893	Aae07893 Modified
15	1800	74.9	1032	4	AAE07901	Aae07901 C. botuli
16	1800	74.9	1059	3	AAY93309	Aay93309 A mangane
17	1800	74.9	1084	3	AAY93312	Aay93312 A mangane
18	1800	74.9	1092	4	AAE07900	Aae07900 C. botuli
19	1783	74.2	432	3	AAY77138	Aay77138 Synthetic
20	1450.5	60.3	452	2	AAW68396	Aaw68396 Clostridi
21	1435	59.7	451	2	AAW68395	Aaw68395 Clostridi
22	1305.5	54.3	419	4	AAB04095	Aab04095 Botulism
23	1305.5	54.3	449	3	AAY77137	Aay77137 Synthetic
24	1305.5	54.3	449	4	AAB04094	Aab04094 Botulism
25	1173.5	48.8	462	2	AAR95009	Aar95009 Type A ne
26	1173.5	48.8	462	2	AAW68390	Aaw68390 Clostridi
27	1094.5	45.5	660	4	AAE07898	Aae07898 Modified
28	1051.5	43.7	445	2	AAW68391	Aaw68391 Clostridi
29	1019.5	42.4	438	2	AAR95008	Aar95008 Type A ne
30	1019.5	42.4	438	2	AAW68389	Aaw68389 Clostridi
31	1019.5	42.4	438	3	AAY77134	Aay77134 Synthetic
32	1013.5	42.2	437	4	AAB04088	Aab04088 Botulism
33	1013.5	42.2	1067	3	AAY93307	Aay93307 A mangane
34	1013.5	42.2	1092	3	AAY93310	Aay93310 A mangane
35	1013.5	42.2	1296	2	AAR95010	Aar95010 C. botuli
36	1011.5	42.1	434	4	AAB04089	Aab04089 Botulism
37	1011.5	42.1	435	4	AAB04090	Aab04090 Botulism
38	1000.5	41.6	432	3	AAY77142	Aay77142 Native bo
39	999.5	41.6	837	3	AAY77140	Aay77140 Native bo
40	992.5	41.3	1295	5	AAU99339	Aau99339 Clostridi
41	988.5	41.1	847	4	AAB04081	Aab04081 Botulism
42	979.5	40.7	415	4	AAB04083	Aab04083 Botulism
43	918.5	38.2	382	3	AAB36303	Aab36303 BoNT/A pr
44	899	37.4	472	2	AAW68393	Aaw68393 Clostridi
45	887	36.9	472	2	AAW68394	Aaw68394 Clostridi

ALIGNMENTS

RESULT 1
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30593.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC express BotC soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which

```
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 448 AA;

  Query Match           100.0%;  Score 2404;  DB 2;  Length 448;
  Best Local Similarity  100.0%;  Pred. No. 6e-182;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 2
AAE07890
ID   AAE07890 standard; protein; 862 AA.
XX
AC   AAE07890;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Geobacillus stearothermophilus.
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Synthetic.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;  99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 9; Page 40; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC   conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC   Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC   translocation domain from diphtheria neurotoxin and a neuronal cell-
CC   specific binding domain from botulinum neurotoxin type F (BoNT/F)
XX
SQ   Sequence 862 AA;

  Query Match            74.9%;  Score 1800.5;  DB 4;  Length 862;
  Best Local Similarity  75.2%;  Pred. No. 9.1e-134;
  Matches  351;  Conservative   39;  Mismatches   52;  Indels   25;  Gaps    6;

Qy          4 HHHHHHHHHSSGH-----IEGRHMASM-----------AILIIYFNRLYKKIKDSSILD 46
              |: ::   :| ||     :   :  |              |||:|||:|||||||:||||
Db        398 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 457

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106
              ||||||||||||||||||||:|||||||||||||||:|: |||||||||||||| ||||
Db        458 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 517

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166
              ||||||||||::  :| | |||||:|: ||||||||||   :   ||||||||:|| :
Db        518 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 574

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226
              |:| | ::  ||:||||||||||||||||||||||||| ||||||||||||||||||
Db        575 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 634

Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286
              ||||:|  ||||||||||:||| |||||||||:||||||::|||||||||:||| |||
Db        635 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 694

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345
              | || || ||  |||||||||  : ::| | :|| ||||||||||  ||||||||||||
Db        695 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 754

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401
              |||||||| ||||||||    | ||  |:||||| |:||||||||||||||||||||||
Db        755 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 814

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              || ||||||||||||||||||:|||||||||| ||||:||:|
Db        815 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 861
```

```
RESULT 3
AAE07892
ID   AAE07892 standard; protein; 887 AA.
XX
AC   AAE07892;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Homo sapiens.
OS   Geobacillus stearothermophilus.
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Synthetic.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;   99GB-00028530.
PR   07-APR-2000;  2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 9; Page 42; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC   conjugate comprises a mitochondrial leader sequence from human Mn-
CC   superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC   linker that can be cleaved by thrombin, translocation domain from
CC   diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC   botulinum neurotoxin type F (BoNT/F)
XX
SQ   Sequence 887 AA;

  Query Match             74.9%;  Score 1800.5;  DB 4;  Length 887;
  Best Local Similarity   75.2%;  Pred. No. 9.5e-134;
  Matches  351;  Conservative   39;  Mismatches   52;  Indels   25;  Gaps    6;

Qy          4 HHHHHHHHHSSGH-----IEGRHMASM-----------AILIIYFNRLYKKIKDSSILD 46
              |: ::   :| ||     :   : |              |||:|||:|||||||:||||
Db        423 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 482

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106
              ||||||||||||||||||||:|||||||||||||:|: ||||||||||||| ||||
Db        483 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 542

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166
              ||||||||||::  :| | |||||:|: ||||||||||   :   |||||||:|| :
Db        543 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 599

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226
              |:| | ::  ||:||||||||||||||||||||||||||| ||||||||||||||||||
Db        600 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 659

Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286
              ||||:|  |||||||||:||| |||||||||:||||||||::|||||||||:||| |||
Db        660 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 719

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345
              | || || ||  |||||||||  : ::| | :|| ||||||||||  ||||||||||||
Db        720 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 779

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401
              |||||||| ||||||||     | ||  |:||||| |:||||||||||||||||||||
Db        780 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 839

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              || ||||||||||||||||||:|||||||||| ||||:||:|
Db        840 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 886


RESULT 4
AAE35713
ID   AAE35713 standard; protein; 979 AA.
XX
AC   AAE35713;
XX
DT   17-JUN-2003  (first entry)
XX
DE   BoNT/F-Hc-DipT HN domain-factor Xa linker-YopT protein fusion construct.
XX
KW   Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW   infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW   muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW   torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW   BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW   botulinum type F neurotoxin; targetted effector protien; YopT.
XX
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Yersinia pestis.
OS   Unidentified.
OS   Chimeric.
XX
PN   WO200296467-A2.
XX
PD   05-DEC-2002.
XX
PF   21-MAY-2002; 2002WO-GB002384.
XX
PR   24-MAY-2001; 2001GB-00012687.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Sutton JM,  Shone CC;
XX
DR   WPI; 2003-167247/16.
XX
PT   Conjugate for modulating cell survival and cell growth, modulating
PT   release of inflammatory mediator from cells, comprises injected bacterial
PT   effector protein and a carrier that targets the protein to target cell.
XX
PS   Example 12; Page 110-114; 130pp; English.
XX
CC   The invention relates to a conjugate comprising an injected bacterial
```

CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and factor Xa linker peptide and
CC Yersinia pestis targetted effector protien YopT. This sequence is used in
CC the exemplification of the invention
XX
SQ Sequence 979 AA;

Query Match 74.9%; Score 1800.5; DB 6; Length 979;
Best Local Similarity 75.2%; Pred. No. 1.1e-133;
Matches 351; Conservative 39; Mismatches 52; Indels 25; Gaps 6;

```
Qy          4 HHHHHHHHHSSGH-----IEGRHMASM-----------AILIIYFNRLYKKIKDSSILD 46
              |: ::   :| ||     :   : |              |||:|||:|||||||:||||
Db        515 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 574

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106
              |||||||||||||||||||||:|||||||||||||:|: ||||||||||||| ||||
Db        575 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 634

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166
              |||||||||::  :| | |||||:|: |||||||||   :   ||||||||:|| :
Db        635 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 691

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226
              |:| | ::  ||:|||||||||||||||||||||||| ||||||||||||||||||||
Db        692 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 751

Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286
              ||||:|  ||||||||||:||| |||||||||:||||||||::|||||||||:||| |||
Db        752 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 811

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345
              | || || ||  |||||||||  : ::| | :|| ||||||||||  ||||||||||||
Db        812 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 871

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401
              |||||||| ||||||||    | ||  |:||||| |:||||||||||||||||||||||
Db        872 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 931

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              || ||||||||||||||||||||:|||||||||| ||||:||:|
Db        932 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 978
```

RESULT 5
AAE35711
ID AAE35711 standard; protein; 1192 AA.
XX
AC AAE35711;
XX
DT 17-JUN-2003 (first entry)
XX
DE BoNT/F-Hc-DipT HN domain-factor Xa-linker SigD protein fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW botulinum type F neurotoxin; invasion gene D protein; SigD protein.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Salmonella typhimurium.
OS Unidentified.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 101-106; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and factor Xa linker peptide and
CC Salmonella typhimurium invasion gene D protein, SigD. This sequence is
CC used in the exemplification of the invention
XX
SQ Sequence 1192 AA;

Query Match 74.9%; Score 1800.5; DB 6; Length 1192;
Best Local Similarity 75.2%; Pred. No. 1.4e-133;
Matches 351; Conservative 39; Mismatches 52; Indels 25; Gaps 6;

```
Qy          4 HHHHHHHHHSSGH-----IEGRHMASM-----------AILIIYFNRLYKKIKDSSILD 46
              |: ::   :| ||     :   : |              |||:|||:|||||||:||||
Db        728 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 787

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106
              |||||||||||||||||||||:|||||||||||||:|: ||||||||||||| ||||
Db        788 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 847

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166
              |||||||||::  :| | |||||:|: |||||||||   :   ||||||||:|| :
Db        848 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 904

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226
              |:| | ::  ||:|||||||||||||||||||||||| ||||||||||||||||||||
Db        905 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 964
```

```
Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286
              ||||:|  |||||||||||:||| |||||||||:||||||||::|||||||||:||| |||
Db        965 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 1024

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345
              | || || ||  |||||||||  : ::| | :|| ||||||||||  ||||||||||||
Db       1025 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 1084

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401
              |||||||| |||||||    | ||  |:||||| |:||||||||||||||||||||||
Db       1085 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 1144

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              || |||||||||||||||||||||:|||||||||| ||||:||:|
Db       1145 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 1191
```

RESULT 6
AAE35710
ID AAE35710 standard; protein; 1192 AA.
XX
AC AAE35710;
XX
DT 17-JUN-2003 (first entry)
XX
DE BoNT/F-Hc-DipT HN domain-thrombin-linker SigD protein fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW botulinum type F neurotoxin; invasion gene D protein; SigD protein.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Salmonella typhimurium.
OS Unidentified.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 97-101; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and thrombin linker peptide and
CC Salmonella typhimurium truncated invasion gene D protein, SigD. This
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 1192 AA;

Query Match 74.9%; Score 1800.5; DB 6; Length 1192;
Best Local Similarity 75.2%; Pred. No. 1.4e-133;
Matches 351; Conservative 39; Mismatches 52; Indels 25; Gaps 6;

```
Qy          4 HHHHHHHHHSSGH-----IEGRHMASM------------AILIIYFNRLYKKIKDSSILD 46
              |: ::   :| ||     :   :  |              |||:|||:|||||||:||||
Db        728 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 787

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106
              |||||||||||||||||||:|||||||||||||:|: ||||||||||||| ||||
Db        788 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 847

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166
              ||||||||||::  :| | |||||:|: ||||||||||   :   |||||||:|| :
Db        848 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 904

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226
              |:| | ::  ||:|||||||||||||||||||||||||| ||||||||||||||||||
Db        905 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 964

Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286
              ||||:|  |||||||||:||| |||||||||:||||||||::|||||||||:||| |||
Db        965 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 1024

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345
              | || || ||  |||||||||  : ::| | :|| ||||||||||  ||||||||||||
Db       1025 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 1084

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401
              |||||||| |||||||    | ||  |:||||| |:||||||||||||||||||||||
Db       1085 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 1144

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              || |||||||||||||||||||||:|||||||||| ||||:||:|
Db       1145 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 1191
```

RESULT 7
AAW09014
ID AAW09014 standard; protein; 431 AA.
XX
AC AAW09014;
XX
DT 17-OCT-2003 (revised)
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum; type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB001409.
XX
PR 12-JUN-1995; 95GB-00011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows recombinant
PT vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX
CC A polypeptide (AAW09014) comprises the heavy chain (amino acids 848-1278)
CC of a type F botulinum neurotoxin (BoNT/F), and can be produced using a
CC synthetic gene (AAT48101) based on the natural gene sequence (AAT48100)
CC for the heavy chain. The polypeptides and its fragments (see also
CC AAW09015-17) lack the light chain and HN epitopes necessary for
CC metalloprotease activity and toxin internalisation. They are free of
CC botulinum toxin activity but can induce protective immunity to a type F
CC botulinum toxin, making them useful for vaccine prodn. Recombinant
CC polypeptides can be produced in transformed host cells, esp. as fusion
CC proteins, e.g. with maltose binding protein to facilitate purification.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 431 AA;

Query Match 74.9%; Score 1800; DB 2; Length 431;
Best Local Similarity 81.0%; Pred. No. 4e-134;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||||:|:
Db          7 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 66

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db         67 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 126

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        127 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 183

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        184 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 243

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        244 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 303

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||| ||||||||    | ||  |:||||| |:||
Db        304 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 363

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db        364 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 423

Qy        442 KENGWKE 448
              ||:||:|
Db        424 KEHGWQE 430
```

RESULT 8
AAB04103
ID AAB04103 standard; protein; 432 AA.
XX
AC AAB04103;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54499.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Fig 18b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 432 AA;

Query Match 74.9%; Score 1800; DB 4; Length 432;
Best Local Similarity 81.0%; Pred. No. 4e-134;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||||:|:
Db          8 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 67

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db         68 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 127

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        128 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 184

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        185 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 244

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
```

```
Db        245 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 304

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||||| ||||||||     | ||  |:||||| |:||
Db        305 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 364

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db        365 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 424

Qy        442 KENGWKE 448
              ||:||:|
Db        425 KEHGWQE 431


RESULT 9
AAB04096
ID   AAB04096 standard; protein; 432 AA.
XX
AC   AAB04096;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54490.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 9b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 432 AA;

  Query Match             74.9%;  Score 1800;  DB 4;  Length 432;
  Best Local Similarity   81.0%;  Pred. No. 4e-134;
  Matches  346;  Conservative   33;  Mismatches   40;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||:|:
Db          8 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 67

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db         68 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 127

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        128 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 184

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||:|  |||||||||:||| |||||||||:|||||||
Db        185 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 244

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        245 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 304

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||||| ||||||||     | ||  |:||||| |:||
Db        305 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 364

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db        365 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 424

Qy        442 KENGWKE 448
              ||:||:|
Db        425 KEHGWQE 431


RESULT 10
AAE07894
ID   AAE07894 standard; protein; 645 AA.
XX
AC   AAE07894;
XX
DT   11-SEP-2003  (revised)
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS   Corynebacterium diphtheriae.
OS   Clostridium botulinum.
OS   Chimeric.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;   99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
```

DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that translocates
PT the therapeutic agent into the neuronal cells.
XX
PS Example 2; Page 44; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin. (Updated on
CC 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 645 AA;

Query Match 74.9%; Score 1800; DB 4; Length 645;
Best Local Similarity 81.0%; Pred. No. 6.8e-134;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||||:|||||||||||||||:|:
Db        221 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 280

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
                |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        281 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 340

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        341 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 397

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        398 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 457

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        458 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 517

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||||| ||||||||     | ||   |:||||| |:||
Db        518 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 577

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db        578 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 637

Qy        442 KENGWKE 448
              ||:||:|
Db        638 KEHGWQE 644
```

RESULT 11
AAE35692
ID AAE35692 standard; protein; 645 AA.
XX
AC AAE35692;
XX
DT 23-OCT-2003 (revised)
DT 17-JUN-2003 (first entry)
XX
DE DipT HN domain-BoNT/F-Hc fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW translocation domain; HN domain; DipT; Hc; botulinum type F neurotoxin;
KW binding domain; BoNT/F.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 57-60; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (Dip-HN domain) and botulinum type F neurotoxin from Clostridium
CC botulinum. This sequence is used in the exemplification of the invention.
CC (Updated on 23-OCT-2003 to standardise OS field)
XX
SQ Sequence 645 AA;

Query Match 74.9%; Score 1800; DB 6; Length 645;
Best Local Similarity 81.0%; Pred. No. 6.8e-134;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||||:|||||||||||||||:|:
Db        221 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 280

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
                |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        281 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 340

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        341 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 397
```

```
Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        398 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 457

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        458 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 517

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||| ||||||||     | ||  |:||||| |:||
Db        518 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 577

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| |||||||||||||||||||:|||||||||| ||
Db        578 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 637

Qy        442 KENGWKE 448
              ||:||:|
Db        638 KEHGWQE 644
```

RESULT 12
AAE35693
ID AAE35693 standard; protein; 657 AA.
XX
AC AAE35693;
XX
DT 17-JUN-2003 (first entry)
XX
DE BoNT/F-Hc-DipT HN domain-thrombin linker fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW botulinum type F neurotoxin.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Unidentified.
OS Chimeric.
XX
PN WO200296467-A2.
XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 60-63; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and thrombin linker peptide. This
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 657 AA;

```
  Query Match             74.9%;  Score 1800;  DB 6;  Length 657;
  Best Local Similarity   81.0%;  Pred. No. 7e-134;
  Matches  346;  Conservative   33;  Mismatches   40;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||:|:
Db        233 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 292

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        293 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 352

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:|||||||||||||||||||||||||||
Db        353 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 409

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        410 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 469

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        470 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 529

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||| ||||||||     | ||  |:||||| |:||
Db        530 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 589

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| |||||||||||||||||||:|||||||||| ||
Db        590 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 649

Qy        442 KENGWKE 448
              ||:||:|
Db        650 KEHGWQE 656
```

RESULT 13
AAE35694
ID AAE35694 standard; protein; 657 AA.
XX
AC AAE35694;
XX
DT 17-JUN-2003 (first entry)
XX
DE BoNT/F-Hc-DipT HN domain-factor Xa linker fusion construct.
XX
KW Apoptosis; therapy; inflammatory mediator; intracellular trafficking;
KW infection; Prion disease; Alzheimer' disease; hypersecretion disorder;
KW muscle spasm; COPD; bronchitis; chronic obstructive pulmonary disease;
KW torticolis; blephorospasm; asthma; fusion protein; diphtheria toxin;
KW BoNT/F; translocation domain; HN domain; DipT; Hc; binding domain;
KW botulinum type F neurotoxin.
XX
OS Corynebacterium diphtheriae.
OS Clostridium botulinum.
OS Unidentified.
OS Chimeric.
XX
PN WO200296467-A2.

XX
PD 05-DEC-2002.
XX
PF 21-MAY-2002; 2002WO-GB002384.
XX
PR 24-MAY-2001; 2001GB-00012687.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Sutton JM, Shone CC;
XX
DR WPI; 2003-167247/16.
XX
PT Conjugate for modulating cell survival and cell growth, modulating
PT release of inflammatory mediator from cells, comprises injected bacterial
PT effector protein and a carrier that targets the protein to target cell.
XX
PS Example 12; Page 63-65; 130pp; English.
XX
CC The invention relates to a conjugate comprising an injected bacterial
CC effector protein and a carrier that targets the effector protein to a
CC target cell. Pharmaceutical composition of the invention is useful for a
CC treatment selected from promoting or inhibiting survival of cells;
CC preventing and reversing damage to cells; killing cells; promoting or
CC inhibiting the growth of cells, apoptosis, release of an inflammatory
CC mediator from cells, division of cells and treating intracellular
CC infection and regulating nitric oxide release from cells. The invention
CC is useful in the manufacture of a medicament for treating a neuronal
CC cell, for intracellular infection, for interfering with intracellular
CC trafficking, for modulating expression of cell-surface markers and for
CC inhibiting secretion from cells. The invention is also useful for
CC treating Prion disease, Alzheimer' disease and wide range of disorders
CC including muscle spasms such as blephorospasm, torticolis and
CC hypersecretion disorders such as chronic obstructive pulmonary disease
CC (COPD), bronchitis and asthma. The present sequence is a fusion construct
CC comprising Corynebacterium diphtheriae diphtheria toxin translocation
CC domain (DipT-HN domain), botulinum type F neurotoxin binding domain
CC (BoNT/F-Hc) from Clostridium botulinum and factor Xa linker peptide. This
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 657 AA;

Query Match 74.9%; Score 1800; DB 6; Length 657;
Best Local Similarity 81.0%; Pred. No. 7e-134;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||||:|:
Db        233 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 292

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        293 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 352

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        353 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 409

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        410 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 469

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        470 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 529

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||| ||||||||     | ||  |:||||| |:||
Db        530 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 589

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||| |||||||||||||||||||:|||||||| ||
Db        590 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 649
```

```
Qy        442 KENGWKE 448
              ||:||:|
Db        650 KEHGWQE 656
```

RESULT 14
AAE07893
ID AAE07893 standard; protein; 685 AA.
XX
AC AAE07893;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Geobacillus stearothermophilus.
OS Influenza virus.
OS Clostridium botulinum.
OS Synthetic.
OS Chimeric.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB004644.
XX
PR 02-DEC-1999; 99GB-00028530.
PR 07-APR-2000; 2000GB-00008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that translocates
PT the therapeutic agent into the neuronal cells.
XX
PS Example 9; Page 43; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation peptide from influenza virus and a neuronal cell-specific
CC binding domain from botulinum neurotoxin type F (BoNT/F)
XX
SQ Sequence 685 AA;

Query Match 74.9%; Score 1800; DB 4; Length 685;
Best Local Similarity 81.0%; Pred. No. 7.4e-134;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
```

```
                ||:|||:|||||||:|||||||||||||||||||||||||||:|||||||||||||||:|:
Db        261 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 320

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        321 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 380

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        381 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 437

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||||:|  |||||||||:||| |||||||||:|||||||
Db        438 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 497

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        498 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 557

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||||| ||||||||    | ||  |:||||| |:||
Db        558 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 617

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db        618 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 677

Qy        442 KENGWKE 448
              ||:||:|
Db        678 KEHGWQE 684


RESULT 15
AAE07901
ID   AAE07901 standard; protein; 1032 AA.
XX
AC   AAE07901;
XX
DT   01-NOV-2001 (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB004644.
XX
PR   02-DEC-1999;  99GB-00028530.
PR   07-APR-2000; 2000GB-00008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that translocates
PT   the therapeutic agent into the neuronal cells.
XX
PS   Example 2; Page 48; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated as
CC   Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention
XX
SQ   Sequence 1032 AA;

  Query Match             74.9%;  Score 1800;  DB 4;  Length 1032;
  Best Local Similarity   81.0%;  Pred. No. 1.3e-133;
  Matches  346;  Conservative   33;  Mismatches   40;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||:|:
Db        608 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 667

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        668 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 727

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                :   ||||||||:|| : |:| | ::  ||:||||||||||||||||||||||||||
Db        728 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 784

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||||:|  |||||||||:||| |||||||||:|||||||
Db        785 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 844

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db        845 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 904

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||||| ||||||||    | ||  |:||||| |:||
Db        905 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 964

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db        965 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 1024

Qy        442 KENGWKE 448
              ||:||:|
Db       1025 KEHGWQE 1031


Search completed: September 10, 2004, 03:33:43
Job time : 31.5516 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 9.5089 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 2404
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1173.5	48.8	462	1	US-08-480-604A-26	Sequence 26, Appl
2	1173.5	48.8	462	2	US-08-405-496A-26	Sequence 26, Appl
3	1173.5	48.8	462	3	US-08-915-136-26	Sequence 26, Appl
4	1173.5	48.8	462	4	US-09-084-517-26	Sequence 26, Appl
5	1019.5	42.4	438	1	US-08-480-604A-23	Sequence 23, Appl
6	1019.5	42.4	438	2	US-08-405-496A-23	Sequence 23, Appl
7	1019.5	42.4	438	3	US-08-915-136-23	Sequence 23, Appl
8	1019.5	42.4	438	4	US-09-084-517-23	Sequence 23, Appl
9	1013.5	42.2	1296	1	US-08-480-604A-28	Sequence 28, Appl
10	1013.5	42.2	1296	2	US-08-405-496A-28	Sequence 28, Appl
11	1013.5	42.2	1296	3	US-08-915-136-28	Sequence 28, Appl
12	1013.5	42.2	1296	4	US-09-084-517-28	Sequence 28, Appl
13	646.5	26.9	1169	4	US-09-255-829-20	Sequence 20, Appl
14	520	21.6	452	1	US-07-618-312A-2	Sequence 2, Appli
15	520	21.6	452	1	US-08-280-228-2	Sequence 2, Appli
16	517	21.5	452	1	US-07-618-312A-4	Sequence 4, Appli
17	517	21.5	452	1	US-08-280-228-4	Sequence 4, Appli
18	517	21.5	618	1	US-08-668-381A-5	Sequence 5, Appli
19	517	21.5	853	4	US-08-913-880C-17	Sequence 17, Appl
20	517	21.5	858	4	US-08-913-880C-16	Sequence 16, Appl
21	517	21.5	860	4	US-08-913-880C-15	Sequence 15, Appl
22	517	21.5	862	4	US-08-913-880C-14	Sequence 14, Appl
23	517	21.5	865	4	US-08-913-880C-13	Sequence 13, Appl
24	517	21.5	866	4	US-08-913-880C-12	Sequence 12, Appl
25	517	21.5	874	4	US-08-913-880C-11	Sequence 11, Appl
26	517	21.5	875	4	US-08-913-880C-10	Sequence 10, Appl
27	517	21.5	1315	4	US-08-913-880C-1	Sequence 1, Appli
28	492	20.5	452	1	US-08-110-786A-8	Sequence 8, Appli
29	264.5	11.0	141	4	US-09-465-276-1	Sequence 1, Appli
30	258.5	10.8	140	3	US-08-446-114A-22	Sequence 22, Appl
31	158	6.6	794	4	US-09-417-485D-8	Sequence 8, Appli
32	155.5	6.5	2184	4	US-09-417-485D-6	Sequence 6, Appli
33	151	6.3	2710	1	US-08-480-604A-6	Sequence 6, Appli
34	151	6.3	2710	2	US-08-405-496A-6	Sequence 6, Appli
35	151	6.3	2710	3	US-08-915-136-6	Sequence 6, Appli
36	151	6.3	2710	4	US-08-957-310-6	Sequence 6, Appli
37	151	6.3	2710	4	US-10-011-366-6	Sequence 6, Appli
38	151	6.3	2710	4	US-09-084-517-6	Sequence 6, Appli
39	150	6.2	23	1	US-08-480-604A-24	Sequence 24, Appl
40	150	6.2	23	2	US-08-405-496A-24	Sequence 24, Appl
41	150	6.2	23	3	US-08-915-136-24	Sequence 24, Appl
42	150	6.2	23	4	US-09-084-517-24	Sequence 24, Appl
43	149	6.2	1004	4	US-09-206-942-57	Sequence 57, Appl
44	149	6.2	1010	4	US-09-206-942-55	Sequence 55, Appl
45	146	6.1	22	4	US-09-660-742-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             48.8%;  Score 1173.5;  DB 1;  Length 462;
  Best Local Similarity   49.4%;  Pred. No. 3.8e-91;
  Matches  230;  Conservative   82;  Mismatches  129;  Indels   25;  Gaps   10;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||| |:  |    | | ::|||::|||:|  ||:| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
               | |:|   |     ::||  ::|   |:: :   | |:||| |:||| |||:||||::
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              :: | ||||||||  ||||||:||      |||||||||   |: ::|:| ::  ||:|
Db        121 SISLNNEYTIINCM-ENNSGWKVSLNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDY 176

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIR 239
              ||:||||||||||| ||:||||| || :| |||||:|| |:||:|: || |   |: |:
Db        177 INRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIK 236

Qy        240 YFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS--- 296
              || :|: ||:: ||: || |: :  |||:||:|| |:| ||: ||   :||: :|:
Db        237 YFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGI 296

Qy        297 -GILNINQQRG-VTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDR 354
               | : :   || |   :::||  || | : ||:|       | || || ||  ||||| :
Db        297 RGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVK 353

Qy        355 GVEYRLYAD-TKSEKEKIIRTSNLND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNI 406
               ||||  : :::  |||:   : |  :| |::|| |      | | | || |:|||::|
Db        354 NKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDI 413

Qy        407 GLLGFHSNN----LVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              | :|||  |    ||||:||   | |:: : || |  |  ::|| |
Db        414 GFIGFHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459


RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match             48.8%;  Score 1173.5;  DB 2;  Length 462;
  Best Local Similarity   49.4%;  Pred. No. 3.8e-91;
  Matches  230;  Conservative   82;  Mismatches  129;  Indels   25;  Gaps   10;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||| |:  |    | | ::|||::|||:|  ||:| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
               | |:|   |     ::||  ::|   |:: :   | |:||| |:||| |||:||||::
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              :: | ||||||||  ||||||:||      |||||||||   |: ::|:| ::  ||:|
Db        121 SISLNNEYTIINCM-ENNSGWKVSLNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDY 176

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIR 239
              ||:||||||||||| ||:||||| || :| |||||:|| |:||:|: || |   |: |:
Db        177 INRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIK 236

Qy        240 YFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS--- 296
              || :|: ||:: ||: || |: :  |||:||:|| |:| ||: ||   :||: :|:
Db        237 YFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGI 296

Qy        297 -GILNINQQRG-VTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDR 354
               | : :   || |   :::||  || | : ||:|       | || || ||  ||||| :
Db        297 RGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVK 353

Qy        355 GVEYRLYAD-TKSEKEKIIRTSNLND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNI 406
               ||||  : :::  |||:    : |  :| |::|| |      | | | || |:|||::|
Db        354 NKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDI 413

Qy        407 GLLGFHSNN----LVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              | :|||  |    ||||:||   | |:: : || |  |  ::|| |
Db        414 GFIGFHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459


RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
```

; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-26

Query Match 48.8%; Score 1173.5; DB 3; Length 462;
Best Local Similarity 49.4%; Pred. No. 3.8e-91;
Matches 230; Conservative 82; Mismatches 129; Indels 25; Gaps 10;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||| |:  |    | | ::|||::|||:|  ||:| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
               | |:|   |     ::|| ::|   |:: :   | |:||| |:||| |||:||||::
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              :: | |||||||  ||||||:||      |||||||||   |: ::|:| ::  ||:|
Db        121 SISLNNEYTIINCM-ENNSGWKVSLNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDY 176

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIR 239
              ||:|||||||||| ||:||||| || :| |||||:|| |:||:||: || |   |: |:
Db        177 INRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIK 236

Qy        240 YFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS--- 296
              || :|: ||:: ||: || |: :  |||::||:|| |:| ||: ||   :||: :|:
Db        237 YFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGI 296

Qy        297 -GILNINQQRG-VTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDR 354
               | : :   || |   :::||  || | : ||:|       | || || ||  ||||| :
Db        297 RGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVK 353

Qy        355 GVEYRLYAD-TKSEKEKIIRTSNLND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNI 406
                |||| : :::  |||:    : | :| |::|| |      | | | || |:|||::|
Db        354 NKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDI 413

Qy        407 GLLGFHSNN----LVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              | :||| |    ||||:||   | |:: : || |  |  ::|| |
Db        414 GFIGFHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459
```

RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-26

```
  Query Match             48.8%;  Score 1173.5;  DB 4;  Length 462;
  Best Local Similarity   49.4%;  Pred. No. 3.8e-91;
  Matches  230;  Conservative   82;  Mismatches  129;  Indels   25;  Gaps   10;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              |||||||||||||||||||||||||| |:  |    | | ::|||::|||:|  ||:| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
                | |:|   |     ::||  ::|   |:: :   | |:||| |:||| |||:||||::
Db         61 ASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFN 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
               :: | ||||||||  ||||||:||       |||||||||   |: ::|:| ::  ||:|
Db        121 SISLNNEYTIINCM-ENNSGWKVSLNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDY 176

Qy        181 INKWIPVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIR 239
              ||:|||||||||||| ||:||||| || :| |||||:|| |:||:||: || |    |: |:
Db        177 INRWIPVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIK 236

Qy        240 YFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS--- 296
              || :|: ||:: ||: || |: :  |||::||:|| |:| ||: ||   :||: :|:
Db        237 YFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGI 296

Qy        297 -GILNINQQRG-VTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDR 354
               | : :   || |   :::||  || | : ||:|       | || || ||  ||||| :
Db        297 RGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVK 353

Qy        355 GVEYRLYAD-TKSEKEKIIRTSNLND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNI 406
                ||||  : :::  |||:    : |  :| |::|| |     | | | || |:|||::|
Db        354 NKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDI 413

Qy        407 GLLGFHSNN----LVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
               | :||| |    ||||:||   | |:: : || |  |  ::|| |
Db        414 GFIGFHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGE 459


RESULT 5
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match             42.4%;  Score 1019.5;  DB 1;  Length 438;
  Best Local Similarity   46.6%;  Pred. No. 3.7e-78;
  Matches  206;  Conservative   82;  Mismatches  129;  Indels   25;  Gaps   10;

Qy         25 MAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 84
               || |:  |    | | ::|||::|||:|  ||:| | | |:|    |      ::||  ::|
Db          1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         85 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKIS 144
                 |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:|
Db         61 LESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVS 119

Qy        145 LRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIPVTITNNRLGNSRIYINGN 204
               |      |||||||||   |: ::|:| ::  ||:|||:|||||||||| ||:|||||
Db        120 LNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIPVTITNNRLNNSKIYINGR 176

Qy        205 LIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDP 263
              || :| |||||:|| |:||:||: || |    |: |:|| :|: ||:: ||: || |: :
Db        177 LIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNS 236

Qy        264 SILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKL 318
               |||::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::||  |
Db        237 GILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSL 296

Qy        319 YEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNL 377
               | | : ||:|       | || || ||  ||||| :  ||||  : :::  |||:    :
Db        297 YRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEI 353

Qy        378 ND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNI 426
               |  :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   |
Db        354 PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQI 413

Qy        427 RRNTSSNGCFWSSISKENGWKE 448
               |:: : || |  |  ::|| |
Db        414 ERSSRTLGCSWEFIPVDDGWGE 435


RESULT 6
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match           42.4%;  Score 1019.5;  DB 2;  Length 438;
  Best Local Similarity  46.6%;  Pred. No. 3.7e-78;
  Matches  206;  Conservative   82;  Mismatches  129;  Indels   25;  Gaps   10;

Qy         25 MAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 84
              ||  |:  |    | | ::|||::|||:|  ||:| | | |:|   |     ::||  ::|
Db          1 MARLLSTFTEYIKNIIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy         85 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKIS 144
                 |:: :    | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:|
Db         61 LESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVS 119

Qy        145 LRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGN 204
              |       |||||||||    |: ::|:| ::  ||:|||:|||||||||||| ||:|||||
Db        120 LNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGR 176

Qy        205 LIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDP 263
              || :| |||||:|| |:||:||: || |   |: |:|| :|: ||:: ||: || |: : :
Db        177 LIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNS 236

Qy        264 SILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKL 318
              |||::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::|| |
Db        237 GILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSL 296

Qy        319 YEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNL 377
              | | : ||:|       | || || ||  ||||| :  ||||  : :::  |||:    :
Db        297 YRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEI 353

Qy        378 ND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNI 426
                |  :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   |

Db        354 PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQI 413

Qy        427 RRNTSSNGCFWSSISKENGWKE 448
                |:: : || |  |  ::|| |
Db        414 ERSSRTLGCSWEFIPVDDGWGE 435


RESULT 7
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match           42.4%;  Score 1019.5;  DB 3;  Length 438;
  Best Local Similarity  46.6%;  Pred. No. 3.7e-78;
  Matches  206;  Conservative   82;  Mismatches  129;  Indels   25;  Gaps   10;
```

```
Qy          25 MAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 84
               || |:  |    | | ::|||::|||:|  ||:| | | |:|   |     ::||  ::|
Db           1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy          85 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKIS 144
                 |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:|
Db          61 LESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVS 119

Qy         145 LRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGN 204
               |      |||||||||   |: ::|:| ::  ||:|||:||||||||||| ||:|||||
Db         120 LNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGR 176

Qy         205 LIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDP 263
               || :| |||||:|| |:||:||: || |   |: |:|| :|: ||:: ||: || |: :
Db         177 LIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNS 236

Qy         264 SILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKL 318
               |||::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::||  |
Db         237 GILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSL 296

Qy         319 YEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNL 377
               | | : ||:|       | || || ||  ||||| :  |||  : :::  |||:    :
Db         297 YRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEI 353

Qy         378 ND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNI 426
                 |  :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   |
Db         354 PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQI 413

Qy         427 RRNTSSNGCFWSSISKENGWKE 448
                |:: : || |  |  ::|| |
Db         414 ERSSRTLGCSWEFIPVDDGWGE 435
```

RESULT 8
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-23

Query Match 42.4%; Score 1019.5; DB 4; Length 438;
Best Local Similarity 46.6%; Pred. No. 3.7e-78;
Matches 206; Conservative 82; Mismatches 129; Indels 25; Gaps 10;

```
Qy          25 MAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 84
               || |:  |    | | ::|||::|||:|  ||:| | | |:|   |     ::||  ::|
Db           1 MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFN 60

Qy          85 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKIS 144
                 |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:|
Db          61 LESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVS 119

Qy         145 LRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGN 204
               |      |||||||||   |: ::|:| ::  ||:|||:||||||||||| ||:|||||
Db         120 LNY---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGR 176

Qy         205 LIVEKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDP 263
               || :| |||||:|| |:||:||: || |   |: |:|| :|: ||:: ||: || |: :
Db         177 LIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNS 236

Qy         264 SILKNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKL 318
               |||::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::||  |
Db         237 GILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSL 296

Qy         319 YEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNL 377
               | | : ||:|       | || || ||  ||||| :  |||  : :::  |||:    :
Db         297 YRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEI 353

Qy         378 ND--SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNI 426
                 |  :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   |
Db         354 PDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQI 413

Qy         427 RRNTSSNGCFWSSISKENGWKE 448
                |:: : || |  |  ::|| |
Db         414 ERSSRTLGCSWEFIPVDDGWGE 435
```

RESULT 9
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             42.2%;  Score 1013.5;  DB 1;  Length 1296;
  Best Local Similarity   46.5%;  Pred. No. 5.3e-77;
  Matches  204;  Conservative   82;  Mismatches  128;  Indels   25;  Gaps   10;

Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
              |:  |     | | ::|||::|||:|  ||:| | | |:|    |      ::||  ::|
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
              |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:||
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVSLNY 980

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
                 |||||||||    |: ::|:| ::  ||:|||:|||||||||||| ||:||||| ||
Db        981 ---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLID 1037

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              :| |||||:|| |:|:|:||: || |   |: |:|| :|: ||:: ||: || |: :  ||
Db       1038 QKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1097

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
              |::||:|| |:| ||: ||   :||: :|:   | : :   || |   :::||  || |
Db       1098 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1157

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
              : ||:|       | || || ||  ||||| :  |||| : ::::  |||:    : |
Db       1158 TKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
               :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   | |:
Db       1215 GNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy        430 TSSNGCFWSSISKENGWKE 448
                : : || |  |  ::|| |
Db       1275 SRTLGCSWEFIPVDDGWGE 1293


RESULT 10
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             42.2%;  Score 1013.5;  DB 2;  Length 1296;
  Best Local Similarity   46.5%;  Pred. No. 5.3e-77;
  Matches  204;  Conservative   82;  Mismatches  128;  Indels   25;  Gaps   10;

Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
              |:  |     | | ::|||::|||:|  ||:| | | |:|    |      ::||  ::|
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
              |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:||
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVSLNY 980

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
                 |||||||||    |: ::|:| ::  ||:|||:|||||||||||| ||:||||| ||
Db        981 ---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLID 1037
```

```
Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              :| |||||:|| |:||:||: || |   |: |:|| :|: ||:: ||: || |: :  ||
Db       1038 QKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1097

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
              |::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::||  || |
Db       1098 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1157

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
              : ||:|       | || || ||  ||||| :  |||  : :::  |||:    : |
Db       1158 TKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
              :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   | |:
Db       1215 GNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy        430 TSSNGCFWSSISKENGWKE 448
              : : || |  |  ::|| |
Db       1275 SRTLGCSWEFIPVDDGWGE 1293
```

```
RESULT 11
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             42.2%;  Score 1013.5;  DB 3;  Length 1296;
  Best Local Similarity   46.5%;  Pred. No. 5.3e-77;
  Matches  204;  Conservative   82;  Mismatches  128;  Indels   25;  Gaps   10;

Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
              |: |    | | ::|||::|||:|  ||:| | | |:|   |     ::||  ::|
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
              |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:||
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVSLNY 980

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
                 |||||||||   |: ::|:| ::  ||:|||:|||||||||||| ||:||||| ||
Db        981 ---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLID 1037

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              :| |||||:|| |:||:||: || |   |: |:|| :|: ||:: ||: || |: :  ||
Db       1038 QKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1097

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
              |::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::||  || |
Db       1098 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1157

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
              : ||:|       | || || ||  ||||| :  |||  : :::  |||:    : |
Db       1158 TKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
              :| |::|| |     | | | || |:|||::|| :|||  |    ||||:||   | |:
Db       1215 GNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy        430 TSSNGCFWSSISKENGWKE 448
              : : || |  |  ::|| |
Db       1275 SRTLGCSWEFIPVDDGWGE 1293
```

```
RESULT 12
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
```

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             42.2%;  Score 1013.5;  DB 4;  Length 1296;
  Best Local Similarity   46.5%;  Pred. No. 5.3e-77;
  Matches  204;  Conservative   82;  Mismatches  128;  Indels   25;  Gaps   10;

Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVSLNY 980

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
Db        981 ---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLID 1037

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
Db       1038 QKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1097

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
Db       1098 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1157

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
Db       1158 TKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
Db       1215 GNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy        430 TSSNGCFWSSISKENGWKE 448
Db       1275 SRTLGCSWEFIPVDDGWGE 1293


RESULT 13
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-20

  Query Match             26.9%;  Score 646.5;  DB 4;  Length 1169;
  Best Local Similarity   42.1%;  Pred. No. 5.1e-46;
  Matches  139;  Conservative   67;  Mismatches  109;  Indels   15;  Gaps    9;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
Db        848 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTSSA 905

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143
Db        906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGWKI 964

Qy        144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203
Db        965 SIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 1020

Qy        204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
Db       1021 KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY 1080

Qy        263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY 1140

Qy        320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYI 349
Db       1141 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYL 1169


RESULT 14
US-07-618-312A-2
; Sequence 2, Application US/07618312A
```

```
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             21.6%;  Score 520;  DB 1;  Length 452;
  Best Local Similarity   29.3%;  Pred. No. 6.9e-36;
  Matches  132;  Conservative   99;  Mismatches  159;  Indels   60;  Gaps   17;

Qy         39 IKDSSILDMRYENNKFIDISGYGSNI----------SINGNVYIYSTNRNQFGIYNSRLS 88
              :| |:||::   |:  |||||: |::           |||             : |:  |
Db         18 LKKSTILNLDINNDINIDISGFNSSVITYPDAQLVPGINGKA---------IHLVNNESS 68

Qy         89 EVNIAQNNDIIYNSRYQNFSISFWVRIPK----HYKPMNHNREYTIINCMGNNN----SG 140
              || : :  || ||  : ||::|||:|:||    | :    | ||:||: |  ::    ||
Db         69 EVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTN-EYSIISSMKKHSLSIGSG 127

Qy        141 WKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYI-NKWIFVTITNNRLGNSRI 199
              | :||   :   :||||:|::|    : || :  :: : |: |||:|:||||:|| :: :
Db        128 WSVSL---KGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANL 183

Qy        200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
              |||| |:    |: || |   :||  |:  |::   || |  |::|   |:  ||| ||:
Db        184 YINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYT 243

Qy        259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLN--SGILNINQQRGVTEGSVFLNY 316
              :   : |:::||| | |: :|||  :    | : |   :  : :       | | : : |
Db        244 SYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYY 303

Qy        317 -KLYEGVEVIIRKNGPIDISNTDNFVRKNDL--AYI--NVVDRGVEYRLYADTKSEKEKI 371
               :|| |:: ||::  |   :  |:||:  |    |:  |  :  | |    :  :  ::|
Db        304 RRLYNGLKFIIKRYTP--NNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRI 361

Qy        372 IRTS------NLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNN---------L 416
              :|       |  : : : |    :  :   ::  :::||:| |:           |
Db        362 LRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGNDPNRDIL 421

Qy        417 VASSWYYNNIRRNTSSNGCFWSSISKENGW 446
              :||:||:|:::      || |  :  : ||
Db        422 IASNWYFNHLKDKIL--GCDWYFVPTDEGW 449


RESULT 15
US-08-280-228-2
; Sequence 2, Application US/08280228
; Patent No. 5571694
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. 5571694th Glebe Road
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,228
;      FILING DATE:  25-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/618,312
;      FILING DATE:  27-NOV-1990
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mary J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  117-163
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-280-228-2

  Query Match             21.6%;  Score 520;  DB 1;  Length 452;
  Best Local Similarity   29.3%;  Pred. No. 6.9e-36;
  Matches  132;  Conservative   99;  Mismatches  159;  Indels   60;  Gaps   17;

Qy         39 IKDSSILDMRYENNKFIDISGYGSNI----------SINGNVYIYSTNRNQFGIYNSRLS 88
              :| |:||::   |:  |||||: |::           |||             : |:  |
Db         18 LKKSTILNLDINNDINIDISGFNSSVITYPDAQLVPGINGKA---------IHLVNNESS 68

Qy         89 EVNIAQNNDIIYNSRYQNFSISFWVRIPK----HYKPMNHNREYTIINCMGNNN----SG 140
              || : :  || ||  : ||::|||:|:||    | :    | ||:||: |  ::    ||
```

```
Db         69 EVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTN-EYSIISSMKKHSLSIGSG 127

Qy        141 WKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYI-NKWIFVTITNNRLGNSRI 199

Db        128 WSVSL---KGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANL 183

Qy        200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258

Db        184 YINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYT 243

Qy        259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLN--SGILNINQQRGVTEGSVFLNY 316

Db        244 SYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYY 303

Qy        317 -KLYEGVEVIIRKNGPIDISNTDNFVRKNDL--AYI--NVVDRGVEYRLYADTKSEKEKI 371

Db        304 RRLYNGLKFIIKRYTP--NNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRI 361

Qy        372 IRTS------NLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNN---------L 416

Db        362 LRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGNDPNRDIL 421

Qy        417 VASSWYYNNIRRNTSSNGCFWSSISKENGW 446

Db        422 IASNWYFNHLKDKIL--GCDWYFVPTDEGW 449
```

Search completed: September 10, 2004, 03:47:12
Job time : 11.5089 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 86.7189 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 2404
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2404	100.0	448	12	US-10-272-898-73	Sequence 73, Appl
2	2404	100.0	448	15	US-10-354-774-73	Sequence 73, Appl
3	2404	100.0	448	15	US-10-271-012-73	Sequence 73, Appl
4	2404	100.0	448	16	US-10-729-122-73	Sequence 73, Appl
5	2404	100.0	448	16	US-10-729-039-73	Sequence 73, Appl
6	2241	93.2	1274	12	US-10-272-898-71	Sequence 71, Appl
7	2241	93.2	1274	15	US-10-354-774-71	Sequence 71, Appl
8	2241	93.2	1274	15	US-10-271-012-71	Sequence 71, Appl
9	2241	93.2	1274	15	US-10-452-024-6	Sequence 6, Appli
10	2241	93.2	1274	16	US-10-729-122-71	Sequence 71, Appl
11	2241	93.2	1274	16	US-10-729-039-71	Sequence 71, Appl
12	2079.5	86.5	1280	15	US-10-452-024-162	Sequence 162, App
13	1800.5	74.9	862	14	US-10-130-973A-4	Sequence 4, Appli
14	1800.5	74.9	887	14	US-10-130-973A-6	Sequence 6, Appli
15	1800	74.9	431	8	US-08-981-087A-1	Sequence 1, Appli
16	1800	74.9	432	10	US-09-910-186A-16	Sequence 16, Appl
17	1800	74.9	432	10	US-09-910-186A-34	Sequence 34, Appl
18	1800	74.9	432	15	US-10-452-024-178	Sequence 178, App
19	1800	74.9	645	14	US-10-130-973A-8	Sequence 8, Appli
20	1800	74.9	685	14	US-10-130-973A-7	Sequence 7, Appli
21	1800	74.9	1032	14	US-10-130-973A-15	Sequence 15, Appl
22	1800	74.9	1092	14	US-10-130-973A-14	Sequence 14, Appl
23	1800	74.9	1278	15	US-10-452-024-152	Sequence 152, App
24	1800	74.9	1278	15	US-10-205-516-12	Sequence 12, Appl
25	1800	74.9	1288	15	US-10-205-516-26	Sequence 26, Appl
26	1566	65.1	1268	15	US-10-452-024-156	Sequence 156, App
27	1460.5	60.8	452	12	US-10-272-898-56	Sequence 56, Appl
28	1460.5	60.8	452	15	US-10-354-774-56	Sequence 56, Appl
29	1460.5	60.8	452	15	US-10-271-012-56	Sequence 56, Appl
30	1460.5	60.8	452	16	US-10-729-122-56	Sequence 56, Appl
31	1460.5	60.8	452	16	US-10-729-039-56	Sequence 56, Appl
32	1435	59.7	451	12	US-10-272-898-54	Sequence 54, Appl
33	1435	59.7	451	15	US-10-354-774-54	Sequence 54, Appl
34	1435	59.7	451	15	US-10-271-012-54	Sequence 54, Appl
35	1435	59.7	451	16	US-10-729-122-54	Sequence 54, Appl
36	1435	59.7	451	16	US-10-729-039-54	Sequence 54, Appl
37	1315	54.7	1251	15	US-10-452-024-127	Sequence 127, App
38	1311.5	54.6	449	10	US-09-910-186A-14	Sequence 14, Appl
39	1311.5	54.6	1252	12	US-10-272-898-52	Sequence 52, Appl
40	1311.5	54.6	1252	15	US-10-354-774-52	Sequence 52, Appl
41	1311.5	54.6	1252	15	US-10-271-012-52	Sequence 52, Appl
42	1311.5	54.6	1252	15	US-10-452-024-128	Sequence 128, App
43	1311.5	54.6	1252	16	US-10-729-122-52	Sequence 52, Appl
44	1311.5	54.6	1252	16	US-10-729-039-52	Sequence 52, Appl
45	1302.5	54.2	1252	15	US-10-452-024-125	Sequence 125, App

ALIGNMENTS

```
RESULT 1
US-10-272-898-73
; Sequence 73, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
```

```
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-272-898-73

  Query Match             100.0%;  Score 2404;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-181;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 2
US-10-354-774-73
; Sequence 73, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-354-774-73

  Query Match             100.0%;  Score 2404;  DB 15;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-181;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 3
US-10-271-012-73
; Sequence 73, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-271-012-73

  Query Match           100.0%;  Score 2404;  DB 15;  Length 448;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-181;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 4
US-10-729-122-73
; Sequence 73, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-729-122-73

  Query Match           100.0%;  Score 2404;  DB 16;  Length 448;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-181;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 5
US-10-729-039-73
```

```
; Sequence 73, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 73:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 448 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-10-729-039-73

  Query Match             100.0%;  Score 2404;  DB 16;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-181;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGY 60

Qy         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYK 120

Qy        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNY 180

Qy        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRY 240

Qy        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN 300

Qy        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 INQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL 360

Qy        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASS 420

Qy        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448
              ||||||||||||||||||||||||||||
Db        421 WYYNNIRRNTSSNGCFWSSISKENGWKE 448


RESULT 6
US-10-272-898-71
; Sequence 71, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;    INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1274 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-272-898-71

  Query Match             93.2%;  Score 2241;  DB 12;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-168;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
                 ||
Db       1273 KE 1274


RESULT 7
US-10-354-774-71
; Sequence 71, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA;
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1274 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-354-774-71

  Query Match           93.2%;  Score 2241;  DB 15;  Length 1274;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-168;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
                 ||
Db       1273 KE 1274


RESULT 8
US-10-271-012-71
; Sequence 71, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1274 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-271-012-71

  Query Match           93.2%;  Score 2241;  DB 15;  Length 1274;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-168;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy         387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy         447 KE 448
               ||
Db        1273 KE 1274
```

RESULT 9
US-10-452-024-6
; Sequence 6, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-6

Query Match 93.2%; Score 2241; DB 15; Length 1274;
Best Local Similarity 100.0%; Pred. No. 5.6e-168;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy         387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy         447 KE 448
               ||
Db        1273 KE 1274
```

RESULT 10
US-10-729-122-71
; Sequence 71, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-729-122-71

Query Match 93.2%; Score 2241; DB 16; Length 1274;
Best Local Similarity 100.0%; Pred. No. 5.6e-168;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912
```

```
Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy         387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy         447 KE 448
               ||
Db        1273 KE 1274
```

RESULT 11
US-10-729-039-71
; Sequence 71, Application US/10729039
; Publication No. US20040142455A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,039
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 1996/08/28
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-729-039-71

Query Match 93.2%; Score 2241; DB 16; Length 1274;
Best Local Similarity 100.0%; Pred. No. 5.6e-168;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy         387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy         447 KE 448
               ||
Db        1273 KE 1274
```

RESULT 12
US-10-452-024-162
; Sequence 162, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 162
; LENGTH: 1280
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-162

Query Match 86.5%; Score 2079.5; DB 15; Length 1280;
Best Local Similarity 93.2%; Pred. No. 3.2e-155;
Matches 398; Conservative 11; Mismatches 13; Indels 5; Gaps 3;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||: |
Db         854 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYSGR 913

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               |||||||||||||||||||||||||| ||||:||| ||||||||||||||||||||||
Db         914 LSEVNIAQNNDIIYNSRYQNFSISFWVTIPKHYRPMNRNREYTIINCMGNNNSGWKISLR 973

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
```

```
Db        974 TIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGNSRIYINGNLI 1033

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266

Db       1034 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1093

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEG-SVFLNYKLYEGVEVI 325

Db       1094 KDYWGNYLLYNKKYYLFNLLRKDKYITRNSGILNINQQRGVTGGISVFLNYKLYEGVEVI 1153

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381

Db       1154 IRKNAPIDISNTDNFVRKNDLAYINVVDHGVEYRLYADISITKSEKIIKLIRTSNPNDSL 1213

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441

Db       1214 GQIIVMDSIGNNCTMNFQNNDGSNIGLLGFHSDDLVASSWYYNHIRRNTSSNGCFWSFIS 1273

Qy        442 KENGWKE 448

Db       1274 KEHGWKE 1280


RESULT 13
US-10-130-973A-4
; Sequence 4, Application US/10130973A
; Publication No. US20030147895A1
; GENERAL INFORMATION:
;  APPLICANT: Shone, Clifford
;  APPLICANT:  Sutton, John
;  APPLICANT:  Silman, Nigel
;  TITLE OF INVENTION: Constructs for Delivery of Therapeutic Agents to Neuronal Cells
;  FILE REFERENCE: 1581.0920000
;  CURRENT APPLICATION NUMBER: US/10/130,973A
;  CURRENT FILING DATE:  2002-10-21
;  PRIOR APPLICATION NUMBER: PCT/GB00/04644
;  PRIOR FILING DATE: 2000-12-04
;  PRIOR APPLICATION NUMBER: GB 9928530.6
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: GB 008658.7
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 862
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic construct
US-10-130-973A-4

  Query Match             74.9%;  Score 1800.5;  DB 14;  Length 862;
  Best Local Similarity   75.2%;  Pred. No. 2.1e-133;
  Matches  351;  Conservative   39;  Mismatches   52;  Indels   25;  Gaps    6;

Qy          4 HHHHHHHHHSSGH-----IEGRHMASM-----------AILIIYFNRLYKKIKDSSILD 46

Db        398 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 457

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106

Db        458 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 517

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166

Db        518 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 574

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226

Db        575 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 634

Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286

Db        635 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 694

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345

Db        695 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 754

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401

Db        755 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 814

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448

Db        815 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 861


RESULT 14
US-10-130-973A-6
; Sequence 6, Application US/10130973A
; Publication No. US20030147895A1
; GENERAL INFORMATION:
;  APPLICANT: Shone, Clifford
;  APPLICANT:  Sutton, John
;  APPLICANT:  Silman, Nigel
;  TITLE OF INVENTION: Constructs for Delivery of Therapeutic Agents to Neuronal Cells
;  FILE REFERENCE: 1581.0920000
;  CURRENT APPLICATION NUMBER: US/10/130,973A
;  CURRENT FILING DATE:  2002-10-21
;  PRIOR APPLICATION NUMBER: PCT/GB00/04644
;  PRIOR FILING DATE: 2000-12-04
;  PRIOR APPLICATION NUMBER: GB 9928530.6
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: GB 008658.7
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 887
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic construct
US-10-130-973A-6

  Query Match             74.9%;  Score 1800.5;  DB 14;  Length 887;
  Best Local Similarity   75.2%;  Pred. No. 2.1e-133;
  Matches  351;  Conservative   39;  Mismatches   52;  Indels   25;  Gaps    6;

Qy          4 HHHHHHHHHSSGH-----IEGRHMASM-----------AILIIYFNRLYKKIKDSSILD 46

Db        423 HNSYNRSAYSPGHKTQPFLHDGYAVSWNTVRSTMSYTNDKILILYFNKLYKKIKDNSILD 482

Qy         47 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 106

Db        483 MRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQN 542

Qy        107 FSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQDTSGNKEN 166

Db        543 FSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNK---IIWTLQDTAGNNQK 599

Qy        167 LIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHVSDNILFK 226

Db        600 LVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFK 659

Qy        227 IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLL 286

Db        660 IVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLL 719

Qy        287 RKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVIIRKNGPIDISNTDNFVRKND 345

Db        720 RTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKND 779

Qy        346 LAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQNN 401
```

```
                ||||||||| ||||||||    | ||  |:||||| |:|||||||||||||||||||||||
Db        780 LAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNN 839

Qy        402 NGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGWKE 448
              || |||||||||||||||||||||:|||||||||| ||||:||:|
Db        840 NGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFISKEHGWQE 886


RESULT 15
US-08-981-087A-1
; Sequence 1, Application US/08981087A
; Publication No. US20020081304A1
;  GENERAL INFORMATION:
;    APPLICANT:  Elmore, Michael J.
;    APPLICANT:  Mauchline, Margaret L.
;    APPLICANT:  Minton, Nigel P.
;    APPLICANT:  Pasechnik, Vladimir A.
;    APPLICANT:  Titball, Richard W.
;    TITLE OF INVENTION:  TYPE F BOTULINUM TOXIN AND USE THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. US20020081304A1th Glebe Rd.  8th floor
;      CITY:  Arlington
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201-4741
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,087A
;      FILING DATE:  27-MAY-1998
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/01409
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9511909.5
;      FILING DATE:  12-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford, Arthur R.
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  124-688
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  431 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-981-087A-1

  Query Match             74.9%;  Score 1800;  DB 8;  Length 431;
  Best Local Similarity   81.0%;  Pred. No. 9.2e-134;
  Matches  346;  Conservative   33;  Mismatches   40;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||:|:
Db          7 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 66

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||| |||||||||||||::  :| | |||||:|: ||||||||||
Db         67 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 126

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               :   ||||||||:|| : |:| | ::  ||:|||||||||||||||||||||||||||
Db        127 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 183

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              |||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db        184 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 243

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||:||| |||| || || ||  |||||||  : ::| | :|| |||||
Db        244 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 303

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||||| ||||||||     | ||  |:||||| |:||
Db        304 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 363

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||| |||||||||||||||||||||:|||||||||| ||
Db        364 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 423

Qy        442 KENGWKE 448
              ||:||:|
Db        424 KEHGWQE 430


Search completed: September 10, 2004, 04:58:18
Job time : 87.7189 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 7.91459 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 2404
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2241	93.2	1274	2	I40813	neurotoxin type F
2	1566	65.1	1268	2	S33411	botulinum neurotox
3	1311.5	54.6	1252	2	S21178	botulinum neurotox
4	1286	53.5	1251	2	JH0256	botulinum neurotox
5	1013.5	42.2	1296	1	BTCLAB	bontoxilysin (EC 3
6	973.5	40.5	1296	2	I40645	botulinum neurotox
7	803	33.4	369	2	S48109	neurotoxin type F
8	749	31.2	1291	2	I40631	non-proteolytic bo
9	735	30.6	1291	1	A48940	bontoxilysin (EC 3
10	663.5	27.6	1297	2	S39791	neurotoxin - Clost
11	612.5	25.5	366	2	S48110	neurotoxin type F
12	536.5	22.3	1276	2	S11455	botulinum neurotox
13	517	21.5	1315	1	BTCLTN	tentoxilysin (EC 3
14	509.5	21.2	1291	2	S46431	botulinum neurotox
15	509.5	21.2	1291	2	A49777	botulinum neurotox
16	497	20.7	367	2	S48106	neurotoxin type E
17	458.5	19.1	1285	2	S70582	botulinum neurotox
18	253	10.5	1162	2	I40817	botulinum toxin no
19	246	10.2	1162	2	A47708	progenitor toxin n
20	246	10.2	1193	2	S68218	botulinum neurotox
21	232	9.7	1193	2	JC4901	nontoxic-nonhemagg
22	218.5	9.1	1165	2	I40644	botulinum neurotox
23	216	9.0	1196	2	JQ1467	toxin, nontoxic co
24	216	9.0	1196	2	S46430	botulinum neurotox
25	202	8.4	122	2	A53878	type E neurotoxin
26	182.5	7.6	1398	2	H71606	hypothetical prote
27	181.5	7.5	1711	2	T18429	hypothetical prote
28	173.5	7.2	807	2	B71605	hypothetical prote
29	170	7.1	2894	2	C64474	hypothetical prote
30	167.5	7.0	836	2	T18460	hypothetical prote
31	167	6.9	2485	1	H71621	serine/threonine-s
32	164.5	6.8	2364	2	I40884	cytotoxin L - Clos
33	162.5	6.8	960	2	S72284	DNA-directed RNA p
34	162.5	6.8	1855	2	S41649	DNA polymerase - m
35	162	6.7	2206	2	G71611	hypothetical prote
36	161.5	6.7	2339	2	A45597	DNA-directed RNA p
37	160.5	6.7	4688	2	F82885	hypothetical prote
38	160	6.7	537	2	A23770	asparagine-rich pr
39	158.5	6.6	336	2	S04682	ribosomal protein
40	157	6.5	3973	2	B71612	hypothetical prote
41	155	6.4	1121	2	F71613	hypothetical prote
42	155	6.4	1532	2	T18438	hypothetical prote
43	153	6.4	1553	2	T18502	hypothetical prote
44	152.5	6.3	1436	2	D71618	hypothetical prote
45	151	6.3	1119	2	T18491	hypothetical prote

ALIGNMENTS

RESULT 1
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 93.2%; Score 2241; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 2.2e-123;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy         207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy         267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy         327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
                ||
Db       1273 KE 1274


RESULT 2
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             65.1%;  Score 1566;  DB 2;  Length 1268;
  Best Local Similarity   70.3%;  Pred. No. 6.4e-84;
  Matches  300;  Conservative   49;  Mismatches   70;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||    | ||:| |||||:|:||||:||| ||||||||||::||||||||||||:||
Db        845 ILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYSSR 904

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||| ||| ||||||||||:||||||||:   | | |||||||| ||||||||||
Db        905 LSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNNLKNLNNEYTIINCMRNNNSGWKISLN 964

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                 ||||||||:|| : |:| | ::  ||:||||| ||||||||||:|::||||||
Db        965 Y---NNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGNLT 1021

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              :||| |||:||| |||||||||:|  ||||||||:|| |||||||||| :||| :||
Db       1022 DQKSILNLGNIHVDDNILFKIVGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDSTIL 1081

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||||| |||: :  :| || |||||:|||: :: ::| | :|| |||||
Db       1082 KDFWGNYLLYNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNIFSNARLYTGVEVI 1141

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKS---EKE-KIIRTSNLNDSL 381
              ||| |  | |||||||||||  ||||||   ||:||||  :    ||  |: | || | :
Db       1142 IRKVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRISNSNYNS 1201

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |:|:|||||:|||||: |||::||||||| |||||||||| ||| || :|||||| ||
Db       1202 NQMIIMDSIGDNCTMNFKTNNGNDIGLLGFHLNNLVASSWYYKNIRNNTRNNGCFWSFIS 1261

Qy        442 KENGWKE 448
              ||:||:|
Db       1262 KEHGWQE 1268


RESULT 3
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the ligh
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match             54.6%;  Score 1311.5;  DB 2;  Length 1252;
  Best Local Similarity   58.7%;  Pred. No. 4.8e-69;
  Matches  250;  Conservative   80;  Mismatches   83;  Indels   13;  Gaps    7;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:|||||| :
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              ||||||:||: |||:::|:|||||||||||| :  | :| | |||||||| :|||||:||
Db        895 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 954

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                   |||||||| :|  : | | |    | ||:|||||||||||||:|||:|::||||||
Db        955 ---NHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1011

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:||||||||||||| |     |:|||||  :|: |||:|||:|||||||: :|
Db       1012 IDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNI 1071
```

```
Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYI-TLNSGILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:|:||| |:|: : :|       |:|| |     :: |  :|| |::|
Db       1072 LKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIR----STILLANRLYSGIKV 1127

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSEKEKIIRTSNLNDSLG 382
              |::    : |  || |||||  ||| |  :   : |||||  : ||| |: |:  :
Db       1128 KIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFN 1185

Qy        383 QIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISK 442
              |::||:|:|||||||:||||:||||||| :: :|||:||| ::| :|:||||||: ||:
Db       1186 QVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISE 1245

Qy        443 ENGWKE 448
              |:||:|
Db       1246 EHGWQE 1251
```

RESULT 4
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match   53.5%; Score 1286; DB 2; Length 1251;
  Best Local Similarity   58.5%; Pred. No. 1.5e-67;
  Matches 249; Conservative   79; Mismatches   84; Indels   14; Gaps   8;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:|||||| :
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              ||||||:||: |||:::|:|||||||||||| :  | :| | |||||||| :||||||:||
Db        895 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 954

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                    |||||||| ||  :  | | |    | ||:|||||||||||||:|||:|::||||||
Db        955 ---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1011

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:||||||||||||| |    |:||||| :|: |||:|||:|||:|||: :|
Db       1012 IDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANI 1071

Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYITLNS-GILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:|:||| |:|: : :|       |:|| |     :: |  :|| |::|
Db       1072 LKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR----STILLANRLYSGIKV 1127

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSEKEKIIRTSNLNDSLG 382
              |::    : |  || |||||  ||| |  :     |||||  : ||| |: |:  :
Db       1128 KIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFN 1185

Qy        383 QIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISK 442
              |::||:|:| ||||||:||||:||||||| :: :|||:||| ::| ||:||| ||: ||:
Db       1186 QVVVMNSVG-NCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISE 1244

Qy        443 ENGWKE 448
              |:||:|
Db       1245 EHGWQE 1250
```

RESULT 5
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025

A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 42.2%; Score 1013.5; DB 1; Length 1296;
Best Local Similarity 46.5%; Pred. No. 1.3e-51;
Matches 204; Conservative 82; Mismatches 128; Indels 25; Gaps 10;

```
Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
              |:  |    | | ::|||::|||:|  ||:| | | |:|    |      ::||  ::|
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
              |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:||
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVSLNY 980

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
                 |||||||||   |: ::|:| ::  ||:|||:||||||||||| ||:|||| ||
Db        981 ---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLID 1037

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              :| |||||:|| |:|||:||: || |   |: |:|| :|: ||:: ||: || |: :  ||
Db       1038 QKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1097

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
              |::||:|| |:| ||: ||   :||: :|:    | : :   || |   :::||  || |
Db       1098 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1157

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
               : ||:|       | || || ||  |||| :  |||  : :::  |||:    : |
Db       1158 TKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
               :| |::|| |      | | | || |:|||::|| :|||   |    ||||:||   : :
Db       1215 GNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1274

Qy        430 TSSNGCFWSSISKENGWKE 448
              : : || |   |  ::|| |
Db       1275 SRTLGCSWEFIPVDDGWGE 1293
```

RESULT 6
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 40.5%; Score 973.5; DB 2; Length 1296;
Best Local Similarity 44.9%; Pred. No. 2.9e-49;
Matches 197; Conservative 84; Mismatches 133; Indels 25; Gaps 10;

```
Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
              |:  |    | | ::|||:|  ||:| | | |:|    |      ::||  ::|  : |
Db        862 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 921

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
              | : :   | |:||| |:||| |||::|||::  :| | ||||||:  ||||||:||
Db        922 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKINLNNEYTIINCI-ENNSGWKVSLNY 980

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
                 ||||||||   | : ::|:| ::  ||:|||:|||||||||||  |:||||| ||
Db        981 ---GEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLID 1037

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              :| |||||:|| |: |:||: || |   |: |:|| :|: ||:: ||: || |: :  ||
Db       1038 QKPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGIL 1097

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
              |::||||| |:| ||: ||   :||: :|:    | : :   || |   :::||  |||| 
Db       1098 KDFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNSTLYEG 1157

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
               : ||:|       | || || ||  |||| :  |||  : :::  |||:    : |
Db       1158 TKFIIKKYAS---GNEDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1214

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
               :| |::|| |      | | | || |:|||::|| :|||   |    ||||:||   : :
Db       1215 GNLSQVVVMKSKDDQGIRNKCKMNLQDNNGNDIGFIGFHLYDNIAKLVASNWYNRQVGKA 1274

Qy        430 TSSNGCFWSSISKENGWKE 448
              : : || |   |  ::|| |
Db       1275 SRTFGCSWEFIPVDDGWGE 1293
```

RESULT 7
S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin

Query Match 33.4%; Score 803; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 5.6e-40;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 279

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        280 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 339

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNR 176
              ||||||||||||||||||||||||||||||
Db        340 TVRDCEIIWTLQDTSGNKENLIFRYEELNR 369


RESULT 8
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match          31.2%;  Score 749;  DB 2;  Length 1291;
  Best Local Similarity  37.3%;  Pred. No. 3.8e-36;
  Matches  170;  Conservative   96;  Mismatches  152;  Indels   38;  Gaps   15;

Qy         23 ASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGI 82
              ::: |||  ||:   :| :: ||::|| :|  ||:||||: : :   | :   ::||| :
Db        844 SNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKL 901

Qy         83 YNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNS 139
              :|  |:: : || :||:|| : :||:|||:||||:      |    ||||||||  |||
Db        902 TSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCM-KNNS 960

Qy        140 GWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRI 199
              |||||:|  |   ||||| | :|  ::: | |      || |||:| ||||||| | |::|
Db        961 GWKISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKI 1016

Qy        200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
              |||| |      | ::|:: |:  | ||: |  | | :: ::|| :|||:|::: |: :|`
Db       1017 YINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYK 1076

Qy        259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNY 316
              :     ||::||| |:|||:||:||    |: || |  :| :  |  :    : | ::||
Db       1077 IQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINY 1136

Qy        317 K-LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIR 373
              : || | : |||:       | |: ||| |  ::::|    |:|:||    | ::||:
Db       1137 RNLYIGEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFL 1195

Qy        374 T--SNLNDSLGQIIVM---DSIGNNCTMNFQNNNGS--NIGLLGFH-----------SN 414
              :  |: |:    | :    :    :| : |: :  |  :|||:| |           :
Db       1196 SIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKD 1255

Qy        415 NLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
                | ||   ::|    |: || |  | |: || |
Db       1256 YFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 9
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the ty
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisio
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
```

A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 30.6%; Score 735; DB 1; Length 1291;
Best Local Similarity 37.0%; Pred. No. 2.5e-35;
Matches 170; Conservative 82; Mismatches 154; Indels 54; Gaps 15;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               |||  ||:   :| :: ||::||::|  ||:||||: : :   | :   ::||| : :|
Db         848 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTSSA 905

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143
                |:: : || :||:|| : :||:|||:||||:      |    ||||||||  |||||||
Db         906 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGWKI 964

Qy         144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203
               |:|  |   ||||| | :|  ::: | |      || |||:| ||||||| | |::|||||
Db         965 SIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 1020

Qy         204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
                |     | :: ::  :  |:||: |  | | :: ::|| :||||| :: ||  |  :
Db        1021 KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY 1080

Qy         263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
               ||::||| |:|||:||:||   |: || |  :| :  |  :    : | ::||: ||
Db        1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY 1140

Qy         320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIRTSNL 377
                | : |||:       | |: ||| |  |::  :   |:|:|     | |:||:
Db        1141 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLF----- 1194

Qy         378 NDSLGQIIVMDSIGN------------NCTMNFQNNNGS--NIGLLGFH---------- 412
                |  |   |  |              :| : |: :  |   |||:| |
Db        1195 ---LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFE 1251

Qy         413 --SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
                :    | ||   ::|       || |  | |: || |
Db        1252 EYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 10
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 27.6%; Score 663.5; DB 2; Length 1297;
Best Local Similarity 34.3%; Pred. No. 3.8e-31;
Matches 157; Conservative 78; Mismatches 174; Indels 49; Gaps 13;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               |||  ||      |  ::|| : |   : || ||||: ::::  :|        || : ||
Db         853 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSE 912

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN-------REYTIINCMGNNNS 139
                 |:    |:  ::|:| : ||||:|||| ||:    |:|        |||||:|:  |:|
Db         913 NSNITAHQSKFVVYDSMFDNFSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KNDS 967

Qy         140 GWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRI 199
               |||:|::  |   ||||| | :   ::: | |   : ||:|||||  :||||:||||: |
Db         968 GWKVSIKGNR---IIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANI 1024

Qy         200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
               ||||:|   : | ||  |: |::| ||:: | | | :| |: | :|  ||: ||: :||
Db        1025 YINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYW 1084

Qy         259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTE-----GSVF 313
                :   : ||::||| | |: :|||||   :: ||   |        | |       :
Db        1085 IQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-----KASMGETAPRTNFNNAA 1139

Qy         314 LNYK-LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVE-YRLYADTKSEK--- 368
               :||: || |:  ||:|       | || ||: |  |:|: :   | ||:|    |::
Db        1140 INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQT 1199

Qy         369 EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQ---NNNGSNIGLLGF----------H 412
               :  :   | : :   :: :       | | |      :   || |           :
Db        1200 QLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTY 1259

Qy         413 SNNLVASSWYYNNIRRNTSS--NGCFWSSISKENGWKE 448
                |    | ||   |  | :    || |  |  : || |
Db        1260 DNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 11
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 25.5%; Score 612.5; DB 2; Length 366;
Best Local Similarity 77.0%; Pred. No. 7.6e-29;
Matches 114; Conservative 17; Mismatches 14; Indels 3; Gaps 1;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
               |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||:|:
Db         220 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 279
```

```
Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
               ||||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db         280 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 339

Qy         147 TVRDCEIIWTLQDTSGNKENLIFRYEEL 174
                 :   ||||||||:|| : |:| | ::
Db         340 YNK---IIWTLQDTAGNNQKLVFNYTQM 364


RESULT 12
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             22.3%;  Score 536.5;  DB 2;  Length 1276;
  Best Local Similarity   30.2%;  Pred. No. 9.9e-24;
  Matches  139;  Conservative   79;  Mismatches  156;  Indels   87;  Gaps   15;

Qy          33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 92
               |  :  | || || :: : |  :| ||| : : :  || : :   | | : :|   :: :
Db         858 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG-DKIIV 916

Qy          93 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 152
                 ||:|:|:: |:| |:|||::| |     | : |||||| :   |||||:  . :|:
Db         917 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-EQNSGWKL---CIRNGN 970

Qy         153 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 212
               | | ||| :   ::||| | |    : | ||| ||||||| :|  ::|||| |   : |
Db         971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy         213 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 271
               :| :: :   |:| |     |:   : || | :|: ||   :|  :|  :   :::|:|||
Db        1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy         272 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 327
               | | :: :||: |    |:||   | :|            | : |    |||  |
Db        1091 NPLKFDTEYYIINDNYIDRYIAPESNVL-----------VLVQYPDRSKLYTG------ 1132

Qy         328 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 387
                 || | :  :   ||  : |   |  : : ||    | |  ||| ::         :
Db        1133 --NPITIKSVSD---KNPYSRILNGDNIILHMLY---NSRKYMIIRDTD--------TIY 1176

Qy         388 DSIGNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS------ 431
               :|  |: |       |:| | | |:  |   |:|: : |    ::: | ||
Db        1177 ATQGGECSQNCVYALKLQSNLG-NYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIY 1235

Qy         432 ----------------------SNGCFWSSISKENGWKE 448
                                    |   ||  ||:: || |
Db        1236 KPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 13
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with bot
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match             21.5%;  Score 517;  DB 1;  Length 1315;
  Best Local Similarity   29.1%;  Pred. No. 1.4e-22;
  Matches  131;  Conservative   99;  Mismatches  160;  Indels   60;  Gaps   17;

Qy          39 IKDSSILDMRYENNKFIDISGYGSNI----------SINGNVYIYSTNRNQFGIYNSRLS 88
```

```
                :| |:||::   |:   ||||: |::          |||             : |:  |
Db        881 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKA---------IHLVNNESS 931

Qy         89 EVNIAQNNDIIYNSRYQNFSISFWVRIPK----HYKPMNHNREYTIINCMGNNN----SG 140
              || : :  || ||  : ||::|||:|:||    | :    | ||:||: |  ::     ||
Db        932 EVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTN-EYSIISSMKKHSLSIGSG 990

Qy        141 WKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYI-NKWIFVTITNNRLGNSRI 199
              | :||   :   :||||:|::|    : || :  :: : |: |||:|:||||:|| :: :
Db        991 WSVSL---KGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANL 1046

Qy        200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
              |||| |:    |: || |   :||  |:  |::   || |  |::|   |:  ||| ||:
Db       1047 YINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYT 1106

Qy        259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLN--SGILNINQQRGVTEGSVFLNY 316
              :   : |:::||| | |: :|||  :   | : |   :  : :       | | : : |
Db       1107 SYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYY 1166

Qy        317 -KLYEGVEVIIRKNGPIDISNTDNFVRKNDL--AYI--NVVDRGVEYRLYADTKSEKEKI 371
               :|| |:: ||::  |   :  :|||:  |    |:  |  :  | |   :  :  ::|
Db       1167 RRLYNGLKFIIKRYTP--NNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRI 1224

Qy        372 IRTS------NLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNN---------L 416
              :|       |   : : : |   :  :   ::  :::||:| |:          |
Db       1225 LRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGNDPNRDIL 1284

Qy        417 VASSWYYNNIRRNTSSNGCFWSSISKENGW 446
              :||:||:|:::     || |  :  : ||
Db       1285 IASNWYFNHLKDKIL--GCDWYFVPTDEGW 1312
```

RESULT 14
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

  Query Match   21.2%; Score 509.5; DB 2; Length 1291;
  Best Local Similarity   31.2%; Pred. No. 3.8e-22;
  Matches 138; Conservative   81; Mismatches 176; Indels   47; Gaps   15;

```
Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQF--GIYNSRLSEV 90
              |  :  | || || ::   |  :| ||| : :|  |:| :       |  |       :|
Db        862 NEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKV 921

Qy         91 NIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRD 150
              : || :|:||| |::|||||:|| |    :::   ||||: :  |||| |    :
Db        922 IVTQNENIVYNSMYESFSISFWIRINKW---VSNLPGYTIIDSV-KNNSGWSIG---IIS 974

Qy        151 CEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKS 210
              :::||:    ::::: | |:  |    | ||| |||:||| :|| :|||| ||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIK 1033

Qy        211 ISNLGDIHVSDNILFK---------IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEP 261
              :  |  |: |  | |:         |    |   : || | :|  |||  :|  |:::
Db       1034 VKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQ 1093

Qy        262 DPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEG 321
              :::|:|||| | |||:||: |:   ::|:  ||  :  | :|          |    ||
Db       1094 YTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN--------NNDFNEG 1145

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL-------YADTKSEKEKII-- 372
              ::||::   |  :  |  ||  |: | ::      | |       |||  | ::
Db       1146 YKIIIKR---IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202

Qy        373 ---RTSNLNDS-LGQIIVMDSIGNNCTMNFQNN-NGSNI-GLLGFHSNNL-VASSWYYNN 425
                 :| ::||: : ||  |::     :  |::| || || |:    :    :  || :|
Db       1203 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1262

Qy        426 IRRNTSSNGCFWSSI-SKENGW 446
                 |   | : | : |    |
Db       1263 YLVPTVKQGNYASLLESTSTHW 1284
```

RESULT 15
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxir
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutinir
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and who
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   21.2%; Score 509.5; DB 2; Length 1291;
  Best Local Similarity   31.2%; Pred. No. 3.8e-22;
  Matches 138; Conservative   81; Mismatches 176; Indels   47; Gaps   15;

```
Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQF--GIYNSRLSEV 90
              |  :  | || || ::   |  :| ||| : :|  |:| :       |  |       :|
Db        862 NEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKV 921

Qy         91 NIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRD 150
              : || :|:||| |::|||||:|| |    :::   ||||: :  |||| |    :
Db        922 IVTQNENIVYNSMYESFSISFWIRINKW---VSNLPGYTIIDSV-KNNSGWSIG---IIS 974
```

```
Qy        151 CEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKS 210
              :::||:    :::::  | |:  |    | ||| |||:||| :|| :|||||  ||
Db        975 NFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIK 1033

Qy        211 ISNLGDIHVSDNILFK---------IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEP 261
              :  |  |: |  | |:         |    |   : || | :|  |||  :|  |:::
Db       1034 VKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQ 1093

Qy        262 DPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEG 321
              :::|:|||| | |||:||: |:   ::|:  ||  :  | :|         |    ||
Db       1094 YTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN--------NNDFNEG 1145

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL-------YADTKSEKEKII-- 372
              ::||::   |  :  |  ||  |: | ::      | |       |||  | ::
Db       1146 YKIIIKR---IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202

Qy        373 ---RTSNLNDS-LGQIIVMDSIGNNCTMNFQNN-NGSNI-GLLGFHSNNL-VASSWYYNN 425
                :| ::||: : ||  |::     :  |::| || || |:    :    :   || :|
Db       1203 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1262

Qy        426 IRRNTSSNGCFWSSI-SKENGW 446
                |   | : | : |    |
Db       1263 YLVPTVKQGNYASLLESTSTHW 1284
```

Search completed: September 10, 2004, 03:44:17
Job time : 9.91459 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 5.46619 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 2404
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2241	93.2	1274	1	BXF_CLOBO	P30996 clostridium
2	1286	53.5	1250	1	BXE_CLOBO	Q00496 clostridium
3	1286	53.5	1250	1	BXE_CLOBU	P30995 clostridium
4	1013.5	42.2	1295	1	BXA1_CLOBO	P10845 clostridium
5	973.5	40.5	1295	1	BXA2_CLOBO	Q45894 clostridium
6	735	30.6	1290	1	BXB_CLOBO	P10844 clostridium
7	663.5	27.6	1296	1	BXG_CLOBO	Q60393 clostridium
8	536.5	22.3	1276	1	BXD_CLOBO	P19321 clostridium
9	517	21.5	1314	1	TETX_CLOTE	P04958 clostridium
10	509.5	21.2	1290	1	BXC1_CLOBO	P18640 clostridium
11	253	10.5	1162	1	BXEN_CLOBU	Q06366 clostridium
12	246	10.2	1162	1	BXEN_CLOBO	P46082 clostridium
13	216	9.0	1196	1	BXCN_CLOBO	P46081 clostridium
14	170	7.1	2894	1	YD96_METJA	Q58791 methanococc
15	161.5	6.7	2339	1	RPC1_PLAFA	P27625 plasmodium
16	160	6.7	537	1	ARP_PLAFA	P04931 plasmodium
17	158.5	6.6	339	1	RMAR_CANGA	P21358 candida gla
18	151	6.3	2710	1	TOXA_CLODI	P16154 clostridium
19	147	6.1	386	1	RMAR_HANWI	P48849 hansenula w
20	146	6.1	970	1	Y087_BUCAI	P57189 buchnera ap
21	144.5	6.0	587	1	DRTS_PLABA	Q27713 plasmodium
22	144.5	6.0	1956	1	ATX1_PLAFA	Q04956 plasmodium
23	143.5	6.0	2136	1	YCF2_MARPO	P09975 marchantia
24	142.5	5.9	1024	1	RPOB_PLAFA	P21421 plasmodium
25	138	5.7	567	1	NU2M_HANWI	P48906 hansenula w
26	134	5.6	817	1	RPOD_ASTLO	P58132 astasia lon
27	134	5.6	850	1	PRSN_PIEBR	Q9gv36 pieris bras
28	134	5.6	1197	1	DPOM_PODAN	Q01529 podospora a
29	133	5.5	1386	1	RPOD_MARPO	P06274 marchantia
30	131	5.4	575	1	RPOC_PLAFA	P21422 plasmodium
31	130.5	5.4	964	1	DPOL_CBEPV	P30319 choristoneu
32	129.5	5.4	396	1	RMAR_YEAST	P02381 saccharomyc
33	128.5	5.3	850	1	PRSN_PIERA	Q9u8q4 pieris rapa
34	128	5.3	986	1	EPIB_STAEP	P30195 staphylococ
35	127.5	5.3	802	1	CSD2_ECOLI	P53513 escherichia
36	125.5	5.2	633	1	ETF1_FOWPV	P21966 fowlpox vir
37	125.5	5.2	819	1	CSD1_ECOLI	P53512 escherichia
38	125	5.2	334	1	NLPD_BUCAI	P57493 buchnera ap
39	125	5.2	709	1	ETF2_FOWPV	Q9j562 fowlpox vir
40	125	5.2	715	1	PTA_BUCBP	Q89as7 buchnera ap
41	125	5.2	968	1	Y080_BUCAP	Q8ka32 buchnera ap
42	124.5	5.2	635	1	ETF1_MYXVL	Q9q819 myxoma viru
43	124	5.2	747	1	Y030_UREPA	Q9prb5 ureaplasma
44	123.5	5.1	544	1	YE72_METJA	Q58867 methanococc
45	123	5.1	558	1	VB04_VACCV	P24769 vaccinia vi

ALIGNMENTS

```
RESULT 1
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
```

```
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             93.2%;  Score 2241;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-127;
  Matches  422;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 912

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 972

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 1032

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1092

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVII 1152

Qy        327 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 RKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIV 1212

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKENGW 1272

Qy        447 KE 448
              ||
Db       1273 KE 1274


RESULT 2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
```

```
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             53.5%;  Score 1286;  DB 1;  Length 1250;
  Best Local Similarity   58.0%;  Pred. No. 5.1e-70;
  Matches  247;  Conservative   80;  Mismatches   85;  Indels   14;  Gaps    8;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:|||||| :
Db        834 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 893

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              ||||||:||: |||:::|:|||||||||||| :  | :| | |||||||| :|||||:||
Db        894 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 953

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                  ||||| :|  |  : | | |   | ||:|||||||||||||:|||:|::||||||
Db        954 ---NHNEIIWTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1010

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:|||||||||||| |    |:|||| :|: |||:|||:|||||||: :|
Db       1011 IDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNI 1070

Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYI-TLNSGILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:|:||| |:|: : :|       |:||  |     :: |  :|| |::|
Db       1071 LKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIR----STILLANRLYSGIKV 1126

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSEKEKIIRTSNLNDSLG 382
               |::    : |  || |||||  ||| |  :   : |||||  : ||| |: |:  :
Db       1127 KIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFN 1184

Qy        383 QIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISK 442
              |::||:|:| ||||||:||||:||||||| :: :|||:||| ::| :|:|||||: ||:
Db       1185 QVVVMNSVG-NCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISE 1243

Qy        443 ENGWKE 448
              |:||:|
Db       1244 EHGWQE 1249


RESULT 3
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
```

```
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             53.5%;  Score 1286;  DB 1;  Length 1250;
  Best Local Similarity   58.5%;  Pred. No. 5.1e-70;
  Matches  249;  Conservative   79;  Mismatches   84;  Indels   14;  Gaps    8;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:|||||| :
Db        834 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 893

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              ||||||:||: |||:::|:||||||||||| :  | :| | |||||||| :||||||:||
Db        894 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 953

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                    |||||||| ||  : | | |    | ||:|||||||||||||:|||:|::||||||
Db        954 ---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1010

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:|||||||||||| |     |:|||| :|: |||:|||:|||:||: :|
Db       1011 IDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANI 1070

Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYITLNS-GILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:|:||| |:|: : :|    :    |:||  |     :: |  :|| |::|
Db       1071 LKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR----STILLANRLYSGIKV 1126

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSEKEKIIRTSNLNDSLG 382
                |::    : |  || |||||  ||| |  :     |||||  : ||| |: |:  :
Db       1127 KIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFN 1184

Qy        383 QIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISK 442
              |::||:|:| |||||:||||:|||||| :: :|||:||| ::| ||:||| ||: ||:
Db       1185 QVVVMNSVG-NCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISE 1243

Qy        443 ENGWKE 448
                |:||:|
Db       1244 EHGWQE 1249


RESULT 4
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
```

```
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             42.2%;  Score 1013.5;  DB 1;  Length 1295;
  Best Local Similarity   46.5%;  Pred. No. 1.3e-53;
  Matches  204;  Conservative   82;  Mismatches  128;  Indels   25;  Gaps   10;

Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
              |:  |     | |  ::|||::|||:|  ||:| | | |:|    |      ::||  ::|
Db        861 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 920

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
              |:: :   | |:||| |:||| |||:||||::  :: | ||||||||  ||||||:||
Db        921 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCM-ENNSGWKVSLNY 979
```

```
Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
Db        980 ---GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLID 1036

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
Db       1037 QKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGIL 1096

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
Db       1097 KDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRG 1156

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
Db       1157 TKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1213

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
Db       1214 GNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERS 1273

Qy        430 TSSNGCFWSSISKENGWKE 448
Db       1274 SRTLGCSWEFIPVDDGWGE 1292
```

```
RESULT 5
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;
```

```
  Query Match             40.5%;  Score 973.5;  DB 1;  Length 1295;
  Best Local Similarity   44.9%;  Pred. No. 3.3e-51;
  Matches  197;  Conservative   84;  Mismatches  133;  Indels   25;  Gaps   10;

Qy         28 LIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRL 87
Db        861 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 920

Qy         88 SEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRT 147
Db        921 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFSKINLNNEYTIINCI-ENNSGWKVSLNY 979

Qy        148 VRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIV 207
Db        980 ---GEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLID 1036

Qy        208 EKSISNLGDIHVSDNILFKIVGC-DDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
Db       1037 QKPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGIL 1096

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNS----GILNINQQRG-VTEGSVFLNYKLYEG 321
Db       1097 KDFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNSTLYEG 1156

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD-TKSEKEKIIRTSNLND- 379
Db       1157 TKFIIKKYAS---GNEDNIVRNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDV 1213

Qy        380 -SLGQIIVMDS-----IGNNCTMNFQNNNGSNIGLLGFHSNN----LVASSWYYNNIRRN 429
Db       1214 GNLSQVVVMKSKDDQGIRNKCKMNLQDNNGNDIGFIGFHLYDNIAKLVASNWYNRQVGKA 1273
```

```
Qy        430 TSSNGCFWSSISKENGWKE 448
              : : || |  |  ::|| |
Db       1274 SRTFGCSWEFIPVDDGWGE 1292


RESULT 6
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             30.6%;  Score 735;  DB 1;  Length 1290;
  Best Local Similarity   37.0%;  Pred. No. 7.3e-37;
  Matches  170;  Conservative   82;  Mismatches  154;  Indels   54;  Gaps   15;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||  ||:   :| :: ||::||::|  ||:||||: : :   | :   ::||| : :|
Db        847 ILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQFKLTSSA 904
```

```
Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143
Db         905 NSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGWKI 963

Qy         144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203
Db         964 SIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYING 1019

Qy         204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
Db        1020 KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY 1079

Qy         263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
Db        1080 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY 1139

Qy         320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIRTSNL 377
Db        1140 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLF----- 1193

Qy         378 NDSLGQIIVMDSIGN------------NCTMNFQNNNGS--NIGLLGFH---------- 412
Db        1194 ---LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFE 1250

Qy         413 --SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
Db        1251 EYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1290
```

RESULT 7
BXG_CLOBO
ID BXG_CLOBO STANDARD; PRT; 1296 AA.
AC Q60393;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE (Bontoxilysin G).
GN BOTG.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;
RX MEDLINE=94092745; PubMed=8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins.";
RL Biochim. Biophys. Acta 1216:487-491(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X74162; CAA52275.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0 0 BY SIMILARITY.
FT CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT CHAIN 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 230 230 BY SIMILARITY.
FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 435 449 INTERCHAIN (PROBABLE).
SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F665C31 CRC64;

Query Match 27.6%; Score 663.5; DB 1; Length 1296;
Best Local Similarity 34.3%; Pred. No. 1.5e-32;
Matches 157; Conservative 78; Mismatches 174; Indels 49; Gaps 13;

```
Qy          27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
Db         852 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSE 911

Qy          87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN-------REYTIINCMGNNNS 139
Db         912 NSNITAHQSKFVVYDSMFDNFSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KNDS 966

Qy         140 GWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRI 199
Db         967 GWKVSIKGNR---IIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANI 1023

Qy         200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
Db        1024 YINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYW 1083

Qy         259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTE-----GSVF 313
Db        1084 IQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-----KASMGETAPRTNFNNAA 1138

Qy         314 LNYK-LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVE-YRLYADTKSEK--- 368
Db        1139 INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQT 1198

Qy         369 EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQ---NNNGSNIGLLGF-------------H 412
Db        1199 QLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTY 1258

Qy         413 SNNLVASSWYYNNIRRNTSS--NGCFWSSISKENGWKE 448
Db        1259 DNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1296
```

RESULT 8
BXD_CLOBO
ID BXD_CLOBO STANDARD; PRT; 1276 AA.
AC P19321;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE (Bontoxilysin D).

```
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             22.3%;  Score 536.5;  DB 1;  Length 1276;
  Best Local Similarity   30.2%;  Pred. No. 6.3e-25;
  Matches  139;  Conservative   79;  Mismatches  156;  Indels   87;  Gaps   15;

Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 92
              |  :  | || || :: : |  :| ||| : : :  || : :   | | : :|   :: :
Db        858 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG-DKIIV 916

Qy         93 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 152
                ||:|:|:: |:| |:|||::| |     | : |||||| :   |||||:    :|:
Db        917 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-EQNSGWKL---CIRNGN 970

Qy        153 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 212
              | | ||| :   ::||| | |    : | ||| ||||||| :|  ::|||| |   : |
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        213 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 271
              :| :: :   |:| |    |:   : || | :|: ||   :|  :|  :   :::|:|||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy        272 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 327
              | | :: :||: |    |:||   | :|             | : |    |||
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVL------------VLVQYPDRSKLYTG------ 1132

Qy        328 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 387
                || | :  :   ||  : |   |  : : ||    | |  ||| ::        :
Db       1133 --NPITIKSVSD---KNPYSRILNGDNIILHMLY---NSRKYMIIRDTD--------TIY 1176

Qy        388 DSIGNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS------ 431
               : |  |: |       |:| | | |:  |   |:|: : |    ::: | ||
Db       1177 ATQGGECSQNCVYALKLQSNLG-NYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIY 1235

Qy        432 ----------------------SNGCFWSSISKENGWKE 448
                                   |   ||  ||:: || |
Db       1236 KPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 9
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
```

```
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
```

```
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190

  Query Match           21.5%;  Score 517;  DB 1;  Length 1314;
  Best Local Similarity  29.1%;  Pred. No. 9.7e-24;
  Matches  131;  Conservative   99;  Mismatches  160;  Indels   60;  Gaps   17;

Qy        39 IKDSSILDMRYENNKFIDISGYGSNI----------SINGNVYIYSTNRNQFGIYNSRLS 88
             :| |:||::   |:   ||||: |::           |||           : |:  |
Db       880 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKA---------IHLVNNESS 930

Qy        89 EVNIAQNNDIIYNSRYQNFSISFWVRIPK----HYKPMNHNREYTIINCMGNNN----SG 140
             || : :  || ||  : ||::|||:|:||    | :    | ||:||: |  ::    ||
Db       931 EVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTN-EYSIISSMKKHSLSIGSG 989

Qy       141 WKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYI-NKWIFVTITNNRLGNSRI 199
             | :||   :   :||||:|::|    : || :  :: : |: |||:|:||||:|| :: :
Db       990 WSVSL---KGNNLIWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANL 1045

Qy       200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
             |||| |:    |: || |   :||  |:  |::   || |  |::|   |:  ||| ||:
Db      1046 YINGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYT 1105

Qy       259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLN--SGILNINQQRGVTEGSVFLNY 316
             :   : |:::||| | |: :|||  :    | : |   :  : :       | | : : |
Db      1106 SYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYY 1165

Qy       317 -KLYEGVEVIIRKNGPIDISNTDNFVRKNDL--AYI--NVVDRGVEYRLYADTKSEKEKI 371
              :|| |:: ||::  |   :  |:||:  |    |:  |  :  | |    :  :  ::|
Db      1166 RRLYNGLKFIIKRYTP--NNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRI 1223

Qy       372 IRTS------NLNDSLGQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNN---------L 416
             :|          |   :  : : |    :  :   ::  :::||:| |:          |
Db      1224 LRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGNDPNRDIL 1283

Qy       417 VASSWYYNNIRRNTSSNGCFWSSISKENGW 446
             :||:||:|:::       || |  :  : ||
Db      1284 IASNWYFNHLKDKIL--GCDWYFVPTDEGW 1311
```

RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO   STANDARD;   PRT; 1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC   RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC   AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC   WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC   INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC   ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC   neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC   detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC   heavy chain (H). The light chain has the pharmacological activity,
CC   while the N- and C-terminal of the heavy chain mediate channel
CC   formation and toxin binding, respectively.

```
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             21.2%;  Score 509.5;  DB 1;  Length 1290;
  Best Local Similarity   31.2%;  Pred. No. 2.7e-23;
  Matches  138;  Conservative   81;  Mismatches  176;  Indels   47;  Gaps   15;

Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQF--GIYNSRLSEV 90
              |  :  | || || ::   |  :| ||| : :|  |:| :      |  |      :|
Db        861 NEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKV 920

Qy         91 NIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRD 150
               : || :|:||| |::||||||:|| |    :::   ||||: :  |||| |    :
Db        921 IVTQNENIVYNSMYESFSISFWIRINKW---VSNLPGYTIIDSV-KNNSGWSIG---IIS 973

Qy        151 CEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKS 210
              :::||:    ::::: | |:  |     | ||| |||:||| :|| :|||| ||
Db        974 NFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIK 1032

Qy        211 ISNLGDIHVSDNILFK---------IVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEP 261
              :  |  |: |  | |:         |   |   : || | :|  |||  :|  |:::
Db       1033 VKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQ 1092

Qy        262 DPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEG 321
               :::|:|||| | |||:||: |:   ::|:  ||  :  | :|          |    ||
Db       1093 YTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRN--------NNDFNEG 1144

Qy        322 VEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRL-------YADTKSEKEKII-- 372
               ::||::    |  :  |  ||  |: | ::       | |        |||  | ::
Db       1145 YKIIIKR---IRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1201

Qy        373 ---RTSNLNDS-LGQIIVMDSIGNNCTMNFQNN-NGSNI-GLLGFHSNNL-VASSWYYNN 425
                :| ::||: : ||  |::      :  |::| || || |:   :   :   || :|
Db       1202 LREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHN 1261

Qy        426 IRRNTSSNGCFWSSI-SKENGW 446
                |   | : | : |    |
Db       1262 YLVPTVKQGNYASLLESTSTHW 1283


RESULT 11
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match             10.5%;  Score 253;  DB 1;  Length 1162;
  Best Local Similarity   27.3%;  Pred. No. 6.4e-08;
  Matches  102;  Conservative   57;  Mismatches  140;  Indels   74;  Gaps   18;

Qy         50 ENNKFIDISG------YGSNI----SINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDII 99
              :||   | ||       |  :|     || |  |: |  ||           ||   ||
Db        811 DNNVIGDTSGKNTLVEYPKDIGLVYGINNNA-IHLTGANQ----------NIKFTNDYF 858

Qy        100 YNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQD 159
               |    |||| ||:|       :| |   : :     :| ||:|      :  ::: : |
Db        859 ENGLTNNFSIYFWLR------NLNQNTIKSKLIGSKEDNCGWEIYF---ENNGLVFNIID 909

Qy        160 TSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNS-RIYINGNLIVEKSISNLGDIH 218
              ::||::|:      |: |||    | :: |: ||| :   |:|:  |:  : |  : :|:
Db        910 SNGNEKNIY-----LSNISN--KSWHYIVISINRLKDQLLIFIDNILVANEDIKEILNIY 962

Qy        219 VSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNK 278
              || |        |   | |    | |  ::  || | | :: : | ::|:    | ||:
Db        963 SSDIISL----LSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNR 1018

Qy        279 KYYLFNLLRKDKYITLNSGILNINQQRGV-----TEGSVFLNYKLYEGVEVIIRKNGPID 333
              | |||      |:     |  | |   :      | |  ||:|  :
Db       1019 TYELFN------YVFPEIAINKIEQNNNIYLSNNNENS--LNFKPLK----------FK 1059
```

```
Qy        334 ISNTD---NFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVMDSI 390
              : ||:    :|:| |    :|:| | |   | |   :  :|    | |:: .  ||  |
Db       1060 LLNTNPNKQYVQKWDEVIFSVLD-GTE--KYLDISIDNNRIQLVDNKNNAKTFIINNDIF 1116

Qy        391 GNNC-TMNFQNNN 402
               :|| |: : | |
Db       1117 ISNCLTLTYNNVN 1129


RESULT 12
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;

  Query Match             10.2%;  Score 246;  DB 1;  Length 1162;
  Best Local Similarity   27.2%;  Pred. No. 1.7e-07;
  Matches  101;  Conservative   60;  Mismatches  139;  Indels   72;  Gaps   18;

Qy         50 ENNKFIDISG------YGSNI----SINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDII 99
              :||   | ||      |  :|     || |  |: |  ||           ||  ||
Db        811 DNNVIGDTSGKNTLVEYPKDIGLVYGINNNA-IHLTGANQ----------NIKFTNDYF 858

Qy        100 YNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCEIIWTLQD 159
               |    |||| ||:|       :  |   : :     :| ||:|      :  ::: : |
Db        859 ENGLTNNFSIYFWLR------NLKQNTIKSKLIGSKEDNCGWEIYF---ENDGLVFNIID 909

Qy        160 TSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNS-RIYINGNLIVEKSISNLGDIH 218
              ::||:|:|:      |: |||    | :: |: ||| :   |:|:  |:  : |  : :|:
Db        910 SNGNEKNIY-----LSNISN--KSWHYIVISINRLKDQLLIFIDNILVANEDIKEILNIY 962

Qy        219 VSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNK 278
               || |        |   | |    | |  ::  || | | :: : | ::|:    | |||:
Db        963 SSDIISL----LSDNNNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNR 1018

Qy        279 KYYLFNLLRKD---KYITLNSGI-LNINQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDI 334
              | ||| :  :     |  |: | |:|| :         ||:|  :             :
```

```
Db       1019 TYELFNYVFPEIAINKIEQNNNIYLSINNENN-------LNFKPLK----------FKL 1060

Qy        335 SNTD---NFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVMDSIG 391
               ||:    :|:| |    :|:| | |   | |  :   :|    | |::   ||  |
Db       1061 LNTNPNKQYVQKWDEVIFSVLD-GTE--KYLDISTTNNRIQLVDNKNNAQIFIINNDIFI 1117

Qy        392 NNC-TMNFQNNN 402
               :|| |: : | |
Db       1118 SNCLTLTYNNVN 1129


RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             9.0%;  Score 216;  DB 1;  Length 1196;
  Best Local Similarity   25.3%;  Pred. No. 1.1e-05;
  Matches   93;  Conservative   66;  Mismatches  142;  Indels   66;  Gaps   19;

Qy         51 NNKFIDISG------YGSNI----SING-NVYIYSTNRNQFGIYNSRLSEVNIAQNNDII 99
              :|   ||||      |  :|     :|| ::|: | |               |  :|
Db        847 SNIIEDISGKNTLIQYTESIELVYGVNGESLYLKSPNE------------TIKFSNKFF 893

Qy        100 YNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGN--NNSGWKISLRTVRDCEIIWTL 157
               |    ||:| ||:|          | : |    :||  || ||:|      |  ::: :
Db        894 TNGLTNNFTICFWLRF------TGKNDDKT--RLIGNKVNNCGWEIYF---EDNGLVFEI 942

Qy        158 QDTSGNKENLIFRYEELNRISNYIN-KWIFVTITNNRLGNS-RIYINGNLIVEKSISNLG 215
              |::||:|::          :|| ||  | :::|: :|| :   |:||   :   ||  :
Db        943 IDSNGNQESVY--------LSNIINDNWYYISISVDRLKDQLLIFINDKNVANVSIDQIL 994

Qy        216 DIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLL 275
              |: | ||:  :|  ::  ||       | :  :    |:   | : | | :::   : |
Db        995 SIY-STNII-SLVNKNNSIYV--EELSVLDNPITSEEVIRNYFSYLDNSYIRDSSKSLLE 1050

Qy        276 YNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVIIRKNGPIDIS 335
```

```
              ||| | |:|      |:   : :  :|         :|: |  :|: :   |   |:|
Db       1051 YNKNYQLYN------YVFPETSLYEVND-----NNKSYLSLKNTDGINISSVKFKLINID 1099

Qy        336 NTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVMDSIGNNCT 395
              :  :|:| |   | |:| | |   | |   | :|   |: :|:  :| |   :
Db       1100 ESKVYVQKWDECIICVLD-GTE--KYLDISPENNRIQLVSS-KDNAKKITVNTDLFRPDC 1155

Qy        396 MNFQNNN 402
              : |  |:
Db       1156 ITFSYND 1162


RESULT 14
YD96_METJA
ID   YD96_METJA     STANDARD;      PRT;  2894 AA.
AC   Q58791;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hypothetical protein MJ1396.
GN   MJ1396.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: Contains 20 PbH1 repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67579; AAB99406.1; -.
DR   PIR; C64474; C64474.
DR   TIGR; MJ1396; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001220; Lectin_legB.
DR   InterPro; IPR006626; PbH1.
DR   Pfam; PF00139; lectin_legB; 1.
DR   SMART; SM00710; PbH1; 20.
KW   Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   REPEAT      543    567       PBH1 1.
FT   REPEAT     2085   2107       PBH1 2.
FT   REPEAT     2135   2156       PBH1 3.
FT   REPEAT     2158   2180       PBH1 4.
FT   REPEAT     2201   2223       PBH1 5.
FT   REPEAT     2224   2244       PBH1 6.
FT   REPEAT     2245   2266       PBH1 7.
FT   REPEAT     2267   2289       PBH1 8.
FT   REPEAT     2290   2311       PBH1 9.
FT   REPEAT     2341   2363       PBH1 10.
FT   REPEAT     2367   2389       PBH1 11.
FT   REPEAT     2390   2419       PBH1 12.
FT   REPEAT     2422   2444       PBH1 13.
FT   REPEAT     2455   2477       PBH1 14.
FT   REPEAT     2479   2501       PBH1 15.
FT   REPEAT     2512   2542       PBH1 16.
FT   REPEAT     2550   2582       PBH1 17.
FT   REPEAT     2589   2611       PBH1 18.
FT   REPEAT     2612   2633       PBH1 19.
FT   REPEAT     2638   2660       PBH1 20.
SQ   SEQUENCE   2894 AA;  322361 MW;  8B64F786537A16DF CRC64;

  Query Match             7.1%;  Score 170;  DB 1;  Length 2894;
  Best Local Similarity   21.0%;  Pred. No. 0.018;
  Matches  121;  Conservative   82;  Mismatches  190;  Indels  184;  Gaps   29;

Qy         29 IIYFNR---LYKKIKDSSILDMRYENNKF----------------IDISGYGSNISINGN 69
              :|| ||   |  :   ||||:    ||                  |  : ||  |  : |
Db       2210 LIYKNRYEGLLLENSSSSILNSNIMNNSIGIYLKENYISKIQKSNISYNAYGIEIVNSSN 2269

Qy         70 VYIYSTNRNQFGIYNSRLSEV------NIAQNNDIIYNSRY--------QNFSI------ 109
              ||| |:|     |:|:    :       |::  | ::||: |          | |:
Db       2270 VYINSSN-----IFNASTDGIAIFNGENVSVENSLLYNNNYSILSYGNLSNLSVLNSLLR 2324

Qy        110 -----SFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKI-SLRTVRDCEIIWTLQDTSGN 163
                  |  : :|     :|: : |   : : : : |  | || :  :  |  :| : |
Db       2325 DSINNSIDIEVPSD-GFLNNLKLYN-SSVLNSGSYGLFIYSLGSASNVNISKSLINGSYK 2382

Qy        164 KENLIFRYEELNRISNYINKWIFVTITNNRL-----GNSRIYINGN-----LIVEKSIS- 212
                |:    :| ::|          |||| |       | | |:||     ||:  :||
Db       2383 DGIYIYGVNAINIVNN--------NITNNGLIGGDPAGSGIKISGNYTKGVLILNNNISH 2434

Qy        213 NLG-------------DIHVSDNILFKIVGCDDET----YVG--IRYFKVFNTELDKTE 252
              |||             |: | :||:    | :: :    |:|  :    :||  :  ::
Db       2435 NLGNGISLEGLWSRTLCDVKVENNIISN-NGIEENSGNGIYIGGRVENVSIFNNTIQYSD 2493

Qy        253 IETLYSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILN----------IN 302
               : :  |       | | ::       : |   :  :|:  | :            |
Db       2494 AQAILIQE------ANGWNSWDWIGTNISIINNTIQYNGLTVTIGNITAGITVGAYGVYN 2547

Qy        303 QQRG--VTEGSVFLNYKL---------YEGVEVIIRKNGPIDIS-NTDNFVRKNDLAYIN 350
              |  |  : ||:  :|  |           |:||       | :  |    :  |: ||
Db       2548 QDNGYIIIEGNKIINNNLCPNPTYGGKVGGIEVYGLNESWISLEFNISKNIIANNSAYGI 2607

Qy        351 VVDRGVEYRLYADTKSEKEK--------------IIRTSNLNDSLGQIIVMDSIGNNCT 395
              ::   : :  :|    ||                | : |  |:||  : :|   :| |
Db       2608 LIGASKDINIINNTIFNNNEKGITIPNWDFVPYNIIISKNSIYNNSL---LGIDLDDDNVT 2664

Qy        396 MN------------------FQNNNGSNIGLLGFHSN--------NLVASSWYYNN--- 425
              :|                  :   || |: : |:  |          | |   :  |
Db       2665 LNDGLLNYNEANHGIDYPIITYAELNGDNLTVKGYIGNGTGSSNFANAVVEIYLVKNLSG 2724

Qy        426 ----IRRNTSSNGCF--------W----SSISKENGW 446
                  | | ||:|          |    | |:  ||:
Db       2725 GDNLIGNNISSDGTVLNDTYGESWIYLGSLIADSNGF 2761


RESULT 15
RPC1_PLAFA
ID   RPC1_PLAFA     STANDARD;      PRT;  2339 AA.
AC   P27625;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92018020; PubMed=1656254;
RA   Li W.B., Bzik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;
RT   "Characterization of the gene encoding the largest subunit of
RT   Plasmodium falciparum RNA polymerase III.";
RL   Mol. Biochem. Parasitol. 46:229-240(1991).
CC   -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
```

```
CC       of DNA into RNA using the four ribonucleoside triphosphates as
CC       substrates.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: RNA polymerase III consists of about 15 different
CC       subunits. This subunit is the largest component of RNA polymerase
CC       III.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are
CC       found in eukaryotic nuclei: polymerase I for the ribosomal RNA
CC       precursor, polymerase II for the mRNA precursor, and polymerase
CC       III for 5S and tRNA genes.
CC   -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73770; AAA29729.1; -.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR007080; RNA_pol_Rpb1_1.
DR   InterPro; IPR007066; RNA_pol_Rpb1_3.
DR   InterPro; IPR007083; RNA_pol_Rpb1_4.
DR   InterPro; IPR007081; RNA_pol_Rpb1_5.
DR   InterPro; IPR006592; RNA_polA_N.
DR   Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR   Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR   Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR   Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR   Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR   SMART; SM00663; RPOLA_N; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW   Zinc-finger; Nuclear protein.
FT   ZN_FING      88    101       C3H-TYPE (POTENTIAL).
SQ   SEQUENCE   2339 AA;  272829 MW;  EDD899363086FD48 CRC64;

  Query Match             6.7%;  Score 161.5;  DB 1;  Length 2339;
  Best Local Similarity   22.2%;  Pred. No. 0.047;
  Matches  120;  Conservative   75;  Mismatches  164;  Indels  181;  Gaps   28;

Qy         29 IIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLS 88
              ||| |        ||   :|   ||   :::   :| |   |::   :||:   | ::  |
Db       1090 IIYNN-------DS---NMNNINNNDSNMNSIHNNNSNMNNIHNNDSNRS---IIHNNDS 1136

Qy         89 EVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNRE---YTIINCMGNNNSGWKISL 145
              :|   |||   || : | |    :      : : || :    :| |   |||: :|
Db       1137 NMNSIHNNDSNMNSIHNNNSNMNNIHNNDSNRSIIHNNDSNMNSIHNNDSNNNNNYK--- 1193

Qy        146 RTVRDC-----------------------------------------EIIWTLQDTSGN 163
                 ||                                         :|: :  |   |
Db       1194 ----DCTHNPYICNESLIIRNIMNRLIYQNIAQEDLFIPLEHDEFLVNKIMESYTDQECN 1249

Qy        164 KENLIFRYEELNRISNYIN----KWIFVT------ITNNRLGNSRIYI--------NGNL 205
              |::| |  :||:  :||:    | : :       || |   |   |:        : |
Db       1250 YEDII-RSLDLNKNVSYIHNDQGKHLSLQMCAEEHITINNTNNDNTYVEQIEMKELSKNK 1308

Qy        206 IVEKSI--SNLGDIH----------VSDNILF-------KIVGCDDE------TYVGIRY 240
               ||      : |:|          ::    |       |:  |:|:       |  |:|
Db       1309 TKEKQSFKGTIRDMHEDSEEQMNKFITKKAKFFIEKKKGKMHECNDDIEYNNTQYDNIQY 1368

Qy        241 FKVFNTELDKTEIE-TLYSNEPDPSILKNYWGNYLLYNKKYY--------------LFNL 285
               :    :    :| | :    | | :||   | :|  |:|:               : ||
Db       1369 NNISCNYIKSQNLENTHHQVNNDLSFIKN---NVILPPKEYHSIFHFVNDYRNVVEIKNL 1425

Qy        286 LRKDKYITLNSGILNINQQRGVTEGSVFLNYKLYEGVEVI--IRKN------GPIDISN 336
              : | | | ||:   |: | :        ::  | : :|:|  | :|           : |
Db       1426 MDKKK-IFLNNSEKNVVQSKYNR-----MSKNLKKKIEIINNIYRNEKKKLNRWKTKMDN 1479

Qy        337 TDNFVRKND---LAYINVVDRGVEYRLYADTKSEKE-------KIIRTSNLNDSLGQIIV 386
                 ||:   :|   :|   :  :  | | |   | |||       |:|  :| ||
Db       1480 DDNYWSSDDDSIIAKKIIKIKNKEKRKY-HPKEEKENFDRNNYKMITDNNNND------- 1531

Qy        387 MDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYN----------NIRRNTSSNGC 435
               ::  ||   |  ||| :|      :|||   :::|||           |   |   | |
Db       1532 -NNNNNNDNNNNDNNNNNN------NSNN---NNYYYNLHDDVNNLGVTNYNTNIYPNDC 1581


Search completed: September 10, 2004, 03:35:53
Job time : 8.46619 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 20.4982 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-73
Perfect score: 2404
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NTSSNGCFWSSISKENGWKE 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2079.5	86.5	1280	2	Q9ZAJ5	Q9zaj5 clostridium
2	1800	74.9	1278	2	Q57236	Q57236 clostridium
3	1566	65.1	1268	2	Q45851	Q45851 clostridium
4	1315	54.7	1251	2	Q9K395	Q9k395 clostridium
5	1302.5	54.2	1252	2	Q8KZM3	Q8kzm3 clostridium
6	1302.5	54.2	1255	2	Q9FAR6	Q9far6 clostridium
7	761	31.7	1291	2	Q8GR96	Q8gr96 clostridium
8	749	31.2	1291	2	Q08077	Q08077 clostridium
9	743	30.9	441	2	Q9X708	Q9x708 clostridium
10	741	30.8	1291	2	Q933K0	Q933k0 clostridium
11	740	30.8	1291	2	Q9ZAJ8	Q9zaj8 clostridium
12	738	30.7	1291	2	Q93G71	Q93g71 clostridium
13	533.5	22.2	1275	12	Q9QTG7	Q9qtg7 clostridium
14	528.5	22.0	1280	2	Q45849	Q45849 clostridium
15	528.5	22.0	1280	2	Q9LBS7	Q9lbs7 clostridium
16	527.5	21.9	1280	2	Q841S3	Q841s3 clostridium
17	518	21.5	451	2	Q9LA13	Q9la13 clostridium
18	509.5	21.2	1291	2	Q93HT3	Q93ht3 clostridium
19	502	20.9	1310	2	Q93N27	Q93n27 clostridium
20	497	20.7	367	2	Q45861	Q45861 clostridium
21	497	20.7	367	2	Q45862	Q45862 clostridium
22	459.5	19.1	1285	2	Q9LBR1	Q9lbr1 clostridium
23	458.5	19.1	1285	2	Q45967	Q45967 clostridium
24	320.5	13.3	361	2	Q45848	Q45848 clostridium
25	320.5	13.3	361	2	Q45846	Q45846 clostridium
26	271	11.3	1198	2	O06018	O06018 clostridium
27	253	10.5	1161	2	Q45891	Q45891 clostridium
28	251	10.4	1162	2	Q9ZAJ6	Q9zaj6 clostridium
29	251	10.4	1193	2	Q45880	Q45880 clostridium
30	251	10.4	1197	2	Q45888	Q45888 clostridium
31	250.5	10.4	1198	2	O69277	O69277 clostridium
32	247.5	10.3	1198	2	Q45893	Q45893 clostridium
33	246	10.2	1193	2	P71107	P71107 clostridium
34	246	10.2	1193	2	Q45914	Q45914 clostridium
35	245	10.2	1161	2	O69276	O69276 clostridium
36	245	10.2	1197	2	Q9ZAJ9	Q9zaj9 clostridium
37	243	10.1	1197	2	P71117	P71117 clostridium
38	243	10.1	1197	2	O33871	O33871 clostridium
39	237.5	9.9	1163	2	Q45850	Q45850 clostridium
40	232	9.7	1193	2	P71108	P71108 clostridium
41	223	9.3	1160	2	O87710	O87710 clostridium
42	222	9.2	1159	2	P71109	P71109 clostridium
43	219	9.1	1196	2	Q53550	Q53550 clostridium
44	219	9.1	1196	9	Q9ZX77	Q9zx77 clostridium
45	218.5	9.1	1165	2	Q45844	Q45844 clostridium

ALIGNMENTS

RESULT 1
Q9ZAJ5
ID Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.

```
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             86.5%;  Score 2079.5;  DB 2;  Length 1280;
  Best Local Similarity   93.2%;  Pred. No. 8.7e-106;
  Matches  398;  Conservative   11;  Mismatches   13;  Indels    5;  Gaps    3;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||: |
Db        854 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYSGR 913

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||||||||||||||||||||||| |||||:||| ||||||||||||||||||||||
Db        914 LSEVNIAQNNDIIYNSRYQNFSISFWVTIPKHYRPMNRNREYTIINCMGNNNSGWKISLR 973

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              |:|||||||||||||||| ||||||||  ||:|||||||||||||||||||||||||||
Db        974 TIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGNSRIYINGNLI 1033

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1034 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 1093

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEG-SVFLNYKLYEGVEVI 325
              |:|||||||||||||||||||||| |||||||||||||| | ||||||||||||||||
Db       1094 KDYWGNYLLYNKKYYLFNLLRKDKYITRNSGILNINQQRGVTGGISVFLNYKLYEGVEVI 1153

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||| |||||||||||||||||||||| |||||||||   |||||  |:||||| ||||
Db       1154 IRKNAPIDISNTDNFVRKNDLAYINVVDHGVEYRLYADISITKSEKIIKLIRTSNPNDSL 1213

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||:||||||||||::|||||||||:|||||||||||||||||||
Db       1214 GQIIVMDSIGNNCTMNFQNNDGSNIGLLGFHSDDLVASSWYYNHIRRNTSSNGCFWSFIS 1273

Qy        442 KENGWKE 448
              ||:||||
Db       1274 KEHGWKE 1280


RESULT 2
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match             74.9%;  Score 1800;  DB 2;  Length 1278;
  Best Local Similarity   81.0%;  Pred. No. 1.7e-90;
  Matches  346;  Conservative   33;  Mismatches   40;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||:|||:|||||||:|||||||||||||||||||||||||||:|||||||||||||:|:
Db        854 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 913

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              ||||||||||||| ||||||||||||||::  :| | |||||:|: ||||||||||
Db        914 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 973

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
              :   ||||||||:|| : |:| | ::   ||:|||||||||||||||||||||||||||
Db        974 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 1030

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
              ||||||||||||||||||||||:|  ||||||||||:||| |||||||||:|||||||
Db       1031 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 1090

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db       1091 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 1150

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYAD---TKSEK-EKIIRTSNLNDSL 381
              |||||  |||||||||||||||||||| ||||||||    | ||  |:||||| |:||
Db       1151 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 1210

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||| ||
Db       1211 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 1270

Qy        442 KENGWKE 448
              ||:||:|
Db       1271 KEHGWQE 1277


RESULT 3
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match           65.1%;  Score 1566;  DB 2;  Length 1268;
  Best Local Similarity  70.3%;  Pred. No. 1.1e-77;
  Matches  300;  Conservative   49;  Mismatches   70;  Indels    8;  Gaps    4;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||    | ||:| |||||:|:||||:||| ||||||||||::||||||||||||:||
Db        845 ILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYSSR 904

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLR 146
              |||||| ||| |||||||||||:||||||||:    | | |||||||| ||||||||||
Db        905 LSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNNLKNLNNEYTIINCMRNNNSGWKISLN 964

Qy        147 TVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLI 206
                    ||||||||:|| : |:| | ::  ||:||||| ||||||||||:|::||||||
Db        965 Y---NNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGNLT 1021

Qy        207 VEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSIL 266
               :||| |||:||| |||||||||:|  ||||||||:|| |||||||||| :||| :||
Db       1022 DQKSILNLGNIHVDDNILFKIVGCNDTRYVGIRYFKIFNMELDKTEIETLYHSEPDSTIL 1081

Qy        267 KNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGV-TEGSVFLNYKLYEGVEVI 325
              |::|||||||||||| |||: :  :| || |||||:|||: :: :: ::| | :|| |||||
Db       1082 KDFWGNYLLYNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNIFSNARLYTGVEVI 1141

Qy        326 IRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKS---EKE-KIIRTSNLNDSL 381
              ||| |  | |||||||||||  ||||||   ||:|||| :   ||  |: | || | :
Db       1142 IRKVGSTDTSNTDNFVRKNDTVYINVVDGNSEYQLYADVSTSAVEKTIKLRRISNSNYNS 1201

Qy        382 GQIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSIS 441
               |:|:|||||:||||||: |||::||||||| |||||||||| ||| || :|||||| ||
Db       1202 NQMIIMDSIGDNCTMNFKTNNGNDIGLLGFHLNNLVASSWYYKNIRNNTRNNGCFWSFIS 1261

Qy        442 KENGWKE 448
              ||:||:|
Db       1262 KEHGWQE 1268


RESULT 4
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
```

```
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match           54.7%;  Score 1315;  DB 2;  Length 1251;
  Best Local Similarity  57.9%;  Pred. No. 5.7e-64;
  Matches  246;  Conservative   84;  Mismatches   83;  Indels   12;  Gaps    6;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:||| ::|| ||:||| |:||:
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFNSK 894

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              |||||:||: |||:::|:||||||||||| :  | :| | |||||||| :||||||:||
Db        895 PSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNINNEYTIINCMRDNNSGWKVSL 954

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                       |||||||| :   : |:|:|   | ||:|||||||||||||:|||:|::||||:|
Db        955 ---NHNEIIWTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGHL 1011

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:||||||||||||| |    |:|||| :|: |||:|||:|||||||: :|
Db       1012 IDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNI 1071

Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYI-TLNSGILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:| ||| |:|: : :|       |:||  |     :: |  :|| |::|
Db       1072 LKDFWGNYLLYDKGYYLLNVLKPNNFIDRRKDSTSLSINNIR----STILLANRLYSGIKV 1127

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTK-SEKEKIIRTSNLNDSLGQ 383
              |::    | |  | ||||||  ||| :     | |||||  ::||| |::|:  :    |
Db       1128 KIQRVN--DSSTNDRFVRKNDQVYINYISNSSSYSLYADTNTTDKEKTIKSSSSGNRFNQ 1185

Qy        384 IIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISKE 443
              ::||:|:||||||||:||||:||||||| :: :|||:||| ::|  :|:||||||: ||:|
```

```
Db       1186 VVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEE 1245

Qy        444 NGWKE 448
                :||:|
Db       1246 HGWQE 1250


RESULT 5
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match           54.2%;  Score 1302.5;  DB 2;  Length 1252;
  Best Local Similarity  58.7%;  Pred. No. 2.7e-63;
  Matches  250;  Conservative   79;  Mismatches   84;  Indels   13;  Gaps    7;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:|||||| :
Db        835 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 894

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              ||||||:||: |||:::|:||||||||||| :  | :| | |||||||| :||||||:||
Db        895 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 954

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                       |||||||| ||  : | | |    | ||:|||||||||||||:|||:|::||||||
Db        955 ---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1011

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:||||||||||||| |    |:|||| :|: |||:|||:|||:|||: :|
Db       1012 IDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANI 1071

Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYITLNS-GILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:|:||| |:|: : :|   :    |:||  |     :: |  :|| |::|
Db       1072 LKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR----STILLANRLYSGIKV 1127

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSEKEKIIRTSNLNDSLG 382
              |::    : |  || |||||  ||| |  :     |||||  : ||| |: |:  :
Db       1128 KIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFN 1185
```

```
Qy        383 QIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISK 442
              |::||:|:||||||||:||||:|||||| :: :|||:||| ::| ||:||| ||: ||:
Db       1186 QVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISE 1245

Qy        443 ENGWKE 448
              |:||:|
Db       1246 EHGWQE 1251


RESULT 6
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match             54.2%;  Score 1302.5;  DB 2;  Length 1255;
  Best Local Similarity   58.7%;  Pred. No. 2.7e-63;
  Matches  250;  Conservative   79;  Mismatches   84;  Indels   13;  Gaps    7;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              ||| |||: :|:|| ||:|:|||:|:|::| ||| |||:|||:|| | ||:|||||| :
Db        838 ILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDK 897

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKH-YKPMNHNREYTIINCMGNNNSGWKISL 145
              ||||||:||: |||:::|:|||||||||||| :  | :| | |||||||| :||||||:||
Db        898 LSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSL 957

Qy        146 RTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNL 205
                    |||||||| ||  : | | |    | ||:|||||||||||:|||:|::||||||
Db        958 ---NHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNL 1014

Qy        206 IVEKSISNLGDIHVSDNILFKIVGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSI 265
              | :||| |||:||||||||||||| |    |:|||| :|: |||:|||:|||:|||: :|
Db       1015 IDKKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANI 1074

Qy        266 LKNYWGNYLLYNKKYYLFNLLRKDKYITLNS-GILNINQQRGVTEGSVFLNYKLYEGVEV 324
              ||::|||||||:|:||| |:|: : :|   :   |:||  |     :: |  :|| |::|
Db       1075 LKDFWGNYLLYDKEYYLLNVLKPNNFINRRTDSTLSINNIR----STILLANRLYSGIKV 1130

Qy        325 IIRKNGPIDISNTDNFVRKNDLAYIN-VVDRGVEYRLYADT-KSEKEKIIRTSNLNDSLG 382
              |::    : |  || |||||  ||| |  :     |||||  : ||| |: |:  :
Db       1131 KIQRVN--NSSTNDNLVRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFN 1188

Qy        383 QIIVMDSIGNNCTMNFQNNNGSNIGLLGFHSNNLVASSWYYNNIRRNTSSNGCFWSSISK 442
              |::||:|:||||||||:||||:|||||| :: :|||:||| ::| ||:||| ||: ||:
Db       1189 QVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDNTNSNGFFWNFISE 1248

Qy        443 ENGWKE 448
              |:||:|
Db       1249 EHGWQE 1254


RESULT 7
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             31.7%;  Score 761;  DB 2;  Length 1291;
  Best Local Similarity   38.4%;  Pred. No. 1.2e-33;
  Matches  174;  Conservative   87;  Mismatches  152;  Indels   40;  Gaps   16;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||  ||:   :| :: ||::|| :|  ||:|||||:|: :   | :   ::||| : :|
Db        848 ILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVEL--NDKNQFKLTSST 905

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143
              ||: : || :||:|| : :||:|||:||||:      |    |||||||:  |||||||
Db        906 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCI-KNNSGWKI 964

Qy        144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203
              |:|  |   ||||| | :|  ::: | |      ||:|||:| |||||||   |::|||||
Db        965 SIRGNR---IIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNS-DNAKIYING 1020

Qy        204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
              |      | ::|::  :  |:||: |  | | :: ::|| :||||| :: |: :|  :
Db       1021 KLESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSY 1080

Qy        263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
              ||:||| |:|||:||:||   |: || |  :| :  |  :   : | ::||: ||
Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLY 1140
```

```
Qy        320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYA--DTKSEKEKIIRTSNL 377
Db       1141 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLF-LANI 1198

Qy        378 NDS-----LGQIIVMDSIGN-NCTMNFQNNNGS--NIGLLGFH-----------SNNLV 417
Db       1199 YDSNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFC 1258

Qy        418 ASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
Db       1259 ISKWYLKEVKRKPYNPNLGCNWQFIPKDEGWIE 1291
```

RESULT 8
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR PIR; I40631; I40631.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 31.2%; Score 749; DB 2; Length 1291;
Best Local Similarity 37.3%; Pred. No. 5.4e-33;
Matches 170; Conservative 96; Mismatches 152; Indels 38; Gaps 15;

```
Qy         23 ASMAILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGI 82
Db        844 SNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKL--NDKNQFKL 901

Qy         83 YNSRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNS 139
Db        902 TSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCM-KNNS 960

Qy        140 GWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRI 199
Db        961 GWKISIRGNR---IIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKI 1016

Qy        200 YINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYS 258
Db       1017 YINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYK 1076

Qy        259 NEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNY 316
Db       1077 IQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINY 1136

Qy        317 K-LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADT--KSEKEKIIR 373
Db       1137 RNLYIGEKFIIRRESNSQSIN-DDIVRKEDYIHLDLVLHHEEWRVYAYKYFKEQEEKLFL 1195

Qy        374 T--SNLNDSLGQIIVM---DSIGNNCTMNFQNNNGS--NIGLLGFH-----------SN 414
Db       1196 SIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYESGVLRKKYKD 1255

Qy        415 NLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
Db       1256 YFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291
```

RESULT 9
Q9X708
ID Q9X708 PRELIMINARY; PRT; 441 AA.
AC Q9X708;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Botulinum neurotoxin type B (Fragment).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99343691; PubMed=10413679;
RA Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA Schiavo G.;
RT "Functional characterisation of tetanus and botulinum neurotoxins
RT binding domains.";
RL J. Cell Sci. 112:2715-2724(1999).
DR EMBL; AJ242628; CAB43706.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 441 441
SQ SEQUENCE 441 AA; 52772 MW; 721D0B468E8C95A4 CRC64;

Query Match 30.9%; Score 743; DB 2; Length 441;
Best Local Similarity 38.4%; Pred. No. 3.6e-33;
Matches 171; Conservative 86; Mismatches 148; Indels 40; Gaps 16;

```
Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
Db          6 ILIEIFNKYNSKILNNIILNLRYRDNNLIDLSGYGANVEVYDGVEL--NDKNQFKLTSST 63

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143
Db         64 NSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCI-KNNSGWKI 122

Qy        144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203
Db        123 SIRGNR---IIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNS-DNAKIYING 178

Qy        204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
Db        179 KLESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSY 238

Qy        263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
```

```
Db        239 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLY 298

Qy        320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYA--DTKSEKEKIIRTSNL 377

Db        299 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNSNREWRVYAYKDFKEEEKKLV-LANI 356

Qy        378 NDS-----LGQIIVMDSIGN-NCTMNFQNNNGS--NIGLLGFH-----------SNNLV 417

Db        357 YDSNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVLKDYKNYFC 416

Qy        418 ASSWYYNNIRRN--TSSNGCFWSSI 440

Db        417 ISKWYLKEVKRKPYNPNLGCNWQFI 441


RESULT 10
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match           30.8%;  Score 741;  DB 2;  Length 1291;
  Best Local Similarity   37.2%;  Pred. No. 1.5e-32;
  Matches  171;  Conservative   81;  Mismatches  154;  Indels   54;  Gaps   15;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86

Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKL--NDKNQFKLTSSA 905

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143

Db        906 NSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGWKI 964

Qy        144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203

Db        965 S---IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS-DNAKIYING 1020

Qy        204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262

Db       1021 KLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSY 1080

Qy        263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319

Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLY 1140

Qy        320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYR--LYADTKSEKEKIIRTSNL 377

Db       1141 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLF----- 1194

Qy        378 NDSLGQIIVMDSIGN-------------NCTMNFQNNNGS--NIGLLGFH---------- 412

Db       1195 ---LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFK 1251

Qy        413 --SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448

Db       1252 EYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 11
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match           30.8%;  Score 740;  DB 2;  Length 1291;
  Best Local Similarity   37.2%;  Pred. No. 1.7e-32;
  Matches  171;  Conservative   81;  Mismatches  154;  Indels   54;  Gaps   15;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86

Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKL--NDKNQFKLTSSA 905

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143

Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGWKI 964

Qy        144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203

Db        965 S---IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS-DNAKIYING 1020
```

```
Qy        204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
              |     | :: ::  :| |:||: |  | | :: ::|| :||||| :: || :|  :
Db       1021 KLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSY 1080

Qy        263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
              ||::||| |:|||:||:||   |: || |  :| :  |  :    : | ::||: ||
Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLY 1140

Qy        320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYR--LYADTKSEKEKIIRTSNL 377
              | : |||:       | |: ||| |  |::  :   |:|  :|   | |:||:
Db       1141 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYMYKYFKKEEEKLF----- 1194

Qy        378 NDSLGQIIVMDSIGN------------NCTMNFQNNNGS--NIGLLGFH---------- 412
              |  |   |   |              :| : |: :  |   |||:| |
Db       1195 ---LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFK 1251

Qy        413 --SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
                :   | ||   ::|    |  || |  | |: || |
Db       1252 EYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 12
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             30.7%;  Score 738;  DB 2;  Length 1291;
  Best Local Similarity   37.2%;  Pred. No. 2.2e-32;
  Matches  171;  Conservative   81;  Mismatches  154;  Indels   54;  Gaps   15;

Qy         27 ILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSR 86
              |||  ||:    | :: ||::|| :|| ||:|||: : :   | :   ::||| : :|
Db        848 ILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKL--NDKNQFKLTSSA 905

Qy         87 LSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN---REYTIINCMGNNNSGWKI 143
              |:: : || :||:|| : :||:|||:||||:      |    ||||||||  |||||||
Db        906 NSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCM-KNNSGWKI 964

Qy        144 SLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYING 203
              |   :|   |||| | :|  ::: | |      || |||:| |||||||   |::|||||
Db        965 S---IRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS-DNAKIYING 1020
```

```
Qy        204 NLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPD 262
              |     | :: ::  :| |:||: |  | | :: ::|| :||||| :: || :|  :
Db       1021 KLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSY 1080

Qy        263 PSILKNYWGNYLLYNKKYYLFNLLRKDKYITL--NSGILNINQQRGVTEGSVFLNYK-LY 319
              ||::||| |:|||:||:||   |: || |  :| :  |  :    : | ::||: ||
Db       1081 SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY 1140

Qy        320 EGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVEYR--LYADTKSEKEKIIRTSNL 377
              | : |||:       | |: ||| |  |::  :   |:|  :|   | |:||:
Db       1141 IGEKFIIRRKSNSQSIN-DDIVRKEDYIYLDFFNLNQEWRVYIYKYFKKEEEKLF----- 1194

Qy        378 NDSLGQIIVMDSIGN------------NCTMNFQNNNGS--NIGLLGFH---------- 412
              |  |   |   |              :| : |: :  |   |||:| |
Db       1195 ---LAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFK 1251

Qy        413 --SNNLVASSWYYNNIRRN--TSSNGCFWSSISKENGWKE 448
                :   | ||   ::|    |  || |  | |: || |
Db       1252 EYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 13
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275    1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match             22.2%;  Score 533.5;  DB 12;  Length 1275;
  Best Local Similarity   30.1%;  Pred. No. 3.3e-21;
  Matches  138;  Conservative   79;  Mismatches  155;  Indels   87;  Gaps   15;

Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 92
              |  :  | || || :: : |  :| ||| : : :  || : :   | | : :|   :: :
Db        858 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSSG-DKIIV 916

Qy         93 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 152
              ||:|:|:: |:| |:|||::| |      | : |||||| :   |||||:    :|:
Db        917 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-EQNSGWKL---CIRNGN 970
```

```
Qy        153 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 212
              | | ||| :   ::||| | |    : | ||| ||||||| :|  ::|||| |   : |
Db        971 IEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIE 1030

Qy        213 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 271
              :| :: :   |:| |    |:   : || | :|: ||   :|  :|  :   :::|:|||
Db       1031 DLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1090

Qy        272 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 327
              | | :: :||: |    |:||   | :|           | : |    ||| |
Db       1091 NPLKFDTEYYIINDNYIDRYIAPESNVL-----------VLVQYPDRSKLYTG------ 1132

Qy        328 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 387
                || | :  :   ||  : |   |  : : ||    | |  ||| ::         :
Db       1133 --NPITIKSVSD---KNPYSRILNGDNIILHMLY---NSRKYMIIRDTD--------TIY 1176

Qy        388 DSIGNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS------ 431
              : |  |: |       |:| | | |:   |   |:|: : |    ::: | ||
Db       1177 ATQGGECSQNCVYALKLQSNLG-NYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIY 1235

Qy        432 ------------------------SNGCFWSSISKENGW 446
                                       |   ||  ||:: ||
Db       1236 KPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGW 1274


RESULT 14
Q45849
ID   Q45849      PRELIMINARY;      PRT;  1280 AA.
AC   Q45849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum C.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=36828;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6813;
RX   MEDLINE=96156810; PubMed=8593068;
RA   Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA   Oguma K.;
RT   "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT   composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT   detection of this gene from Clostridium botulinum type C organisms.";
RL   Appl. Environ. Microbiol. 62:662-667(1996).
DR   EMBL; D49440; BAA08418.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147815 MW;  74F813B228B8C989 CRC64;

  Query Match             22.0%;  Score 528.5;  DB 2;  Length 1280;
  Best Local Similarity   30.3%;  Pred. No. 6.2e-21;
  Matches  141;  Conservative   78;  Mismatches  151;  Indels   95;  Gaps   16;

Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 92
                |  :  | || || :: : |  :| ||| : : : |:| : :   | | : :|   :: :
Db        862 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSSG-DKIIV 920

Qy         93 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 152
                ||:|:|:: |:| |:|||::| |      | : |||||| :   |||||:    :|:
Db        921 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-KQNSGWKL---CIRNGN 974

Qy        153 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 212
              | | ||| :   ::||| | |    : | ||| ||||||| :|  ::|||| |   : |
Db        975 IEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSERIE 1034

Qy        213 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 271
              :| :: :   |:| |    |:   : || | :|: ||   :|  :|  :   :::|:|||
Db       1035 DLNEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1094

Qy        272 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 327
              | | :: :||: |    |:||   | :|           | : |    ||| |
Db       1095 NPLKFDTEYYIINDNYIDRYIAPKSNIL-----------VLVQYPDRSKLYTG------ 1136

Qy        328 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 387
                || | :  :   ||  : |   |  : : ||    | |  ||| :
Db       1137 --NPITIKSVSD---KNPYSRILNGDNIMFHMLY---NSGKYMIIRDT------------ 1176

Qy        388 DSI----GNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS-- 431
              |:|    |  |: |       |:| | | |:  |   |:|: : |    :::  :||
Db       1177 DTIYAIEGRECSKNCVYALKLQSNLG-NYGIGIFSIKNIVSQNKYCSQIFSSFMKNTMLL 1235

Qy        432 ----------------------------SNGCFWSSISKENGWKE 448
                                           |   ||  ||:: || |
Db       1236 ADIYKPWRFSFENAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1280


RESULT 15
Q9LBS7
ID   Q9LBS7      PRELIMINARY;      PRT;  1280 AA.
AC   Q9LBS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-6814;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Kawabe T.,
RA   Murakami F., Nakatsuka M., Ohyama T.;
RT   "Organization of Gene Encoding Components of the Botulinum Progenitor
RT   Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT   Chimeric Sequence in the Gene Encoding Each Component.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037166; BAA89713.1; -.
DR   PIR; A43503; A43503.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147757 MW;  DE124FFB6F68450B CRC64;
```

```
  Query Match             22.0%;  Score 528.5;  DB 2;  Length 1280;
  Best Local Similarity   30.3%;  Pred. No. 6.2e-21;
  Matches  141;  Conservative   78;  Mismatches  151;  Indels   95;  Gaps   16;

Qy         33 NRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNI 92
              |  :  | || || :: : |  :| ||| : : : |:| : :   | | : :|   :: :
Db        862 NEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSSSG-DKIIV 920

Qy         93 AQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHNREYTIINCMGNNNSGWKISLRTVRDCE 152
                ||:|:|:: |:| |:|||::| |     | : |||||| :   |||||:    :|:
Db        921 NLNNNILYSAIYENSSVSFWIKISKDL--TNSHNEYTIINSI-KQNSGWKL---CIRNGN 974

Qy        153 IIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSIS 212
              | | ||| :   ::||| | |    : | ||| ||||||| :|  ::|||| |   : |
Db        975 IEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSERIE 1034

Qy        213 NLGDIHVSDNILFKI-VGCDDETYVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWG 271
              :| :: :   |:| |    |:   : || | :|: ||   :|  :|  :   :::|:|||
Db       1035 DLNEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWG 1094

Qy        272 NYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTEGSVFLNY----KLYEGVEVIIR 327
              | | :: :||: |    |:||   | ||             | : |    ||| |
Db       1095 NPLKFDTEYYIINDNYIDRYIAPKSNIL-----------VLVQYPDRSKLYTG------ 1136

Qy        328 KNGPIDISNTDNFVRKNDLAYINVVDRGVEYRLYADTKSEKEKIIRTSNLNDSLGQIIVM 387
               || | :  :   ||  : |   |  : : ||    | |  ||| :
Db       1137 --NPITIKSVSD---KNPYSRILNGDNIMFHMLY---NSGKYMIIRDT----------- 1176

Qy        388 DSI----GNNCTMN------FQNNNGSNIGLLGFHSNNLVASSWY----YNNIRRNTS-- 431
              |:|    |  |: |       |:| | | |:  |   |:|: : |    :::  :||
Db       1177 DTIYAIEGRECSKNCVYALKLQSNLG-NYGIGIFSIKNIVSQNKYCSQIFSSFMKNTMLL 1235

Qy        432 ---------------------------SNGCFWSSISKENGWKE 448
                                          |   ||  ||:: || |
Db       1236 ADIYKPWRFSFENAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1280


Search completed: September 10, 2004, 03:41:47
Job time : 21.4982 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 75.9935 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 1297
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	728	56.1	848	5	ABG69087	Abg69087 Botulinum
2	448	34.5	449	3	AAY77139	Aay77139 Synthetic
3	448	34.5	449	4	AAB04167	Aab04167 Botulism
4	447	34.5	473	2	AAW68400	Aaw68400 Clostridi
5	441	34.0	444	5	ABG69086	Abg69086 Botulinum
6	408	31.5	410	4	AAB04102	Aab04102 Botulism
7	265	20.4	439	5	ABG69073	Abg69073 Botulinum
8	16	1.2	441	5	ABG69068	Abg69068 Botulinum
9	16	1.2	441	5	ABG69076	Abg69076 Botulinum
10	16	1.2	441	5	ABB80654	Abb80654 Botulinum
11	16	1.2	441	6	ABU61545	Abu61545 Botulinum
12	16	1.2	548	2	AAW56014	Aaw56014 Recombina
13	16	1.2	852	5	ABG69077	Abg69077 Botulinum
14	16	1.2	858	2	AAW56018	Aaw56018 Recombina
15	16	1.2	1169	2	AAW56017	Aaw56017 Recombina
16	16	1.2	1291	2	AAW68392	Aaw68392 Clostridi
17	15	1.2	413	4	AAB04097	Aab04097 Botulism
18	15	1.2	848	4	AAB04082	Aab04082 Botulism
19	15	1.2	1070	3	AAY93308	Aay93308 A mangane
20	15	1.2	1095	3	AAY93311	Aay93311 A mangane
21	14	1.1	1315	4	AAB61169	Aab61169 Clostridi
22	12	0.9	31	2	AAW11975	Aaw11975 T-cell ep
23	12	0.9	31	4	AAM99077	Aam99077 Vaccine r
24	12	0.9	422	5	ABG69071	Abg69071 Botulinum
25	12	0.9	427	5	ABG69082	Abg69082 Botulinum
26	12	0.9	804	5	ABG69083	Abg69083 Botulinum
27	11	0.8	207	4	AAG79295	Aag79295 Amino aci
28	11	0.8	407	3	AAY77141	Aay77141 Native bo
29	11	0.8	413	4	AAB04168	Aab04168 Botulism
30	11	0.8	413	5	ABG69074	Abg69074 Botulinum
31	11	0.8	423	3	AAB36302	Aab36302 C. botuli
32	11	0.8	448	4	AAE07896	Aae07896 Modified
33	11	0.8	452	2	AAR12471	Aar12471 Tetanus t
34	11	0.8	463	2	AAY00921	Aay00921 Tetanus t
35	11	0.8	573	1	AAP70345	Aap70345 Portion o
36	11	0.8	605	4	AAE07897	Aae07897 Modified
37	11	0.8	618	2	AAW48909	Aaw48909 SOD-1/TTC
38	11	0.8	660	4	AAE07898	Aae07898 Modified
39	11	0.8	665	4	AAE07895	Aae07895 Modified
40	11	0.8	665	6	AAE35689	Aae35689 DipT HN d
41	11	0.8	677	6	AAE35690	Aae35690 TeNT-Hc-D
42	11	0.8	677	6	AAE35691	Aae35691 TeNT-Hc-D
43	11	0.8	810	6	AAE35714	Aae35714 TeNT-Hc-D
44	11	0.8	810	6	AAE35715	Aae35715 TeNT-Hc-D
45	11	0.8	837	3	AAY77140	Aay77140 Native bo

ALIGNMENTS

RESULT 1
ABG69087
ID ABG69087 standard; protein; 848 AA.
XX
AC ABG69087;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #21.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98557.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 162-164; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

```
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 848 AA;

  Query Match           56.1%;  Score 728;  DB 5;  Length 848;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  848;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIEN-QLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 599

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 659

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 719

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 779

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 839

Qy        841 IPFDLSLYT 849
              |||||||||
Db        840 IPFDLSLYT 848
```

```
RESULT 2
AAY77139
ID   AAY77139 standard; protein; 449 AA.
XX
AC   AAY77139;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype G (BoNTG) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype G; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87217.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 28; Page 47-48; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 449 AA;

  Query Match           34.5%;  Score 448;  DB 3;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        850 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 61
```

```
Qy        910 NSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGW 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 NSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGW 121

Qy        970 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 181

Qy       1030 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 241

Qy       1090 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 301

Qy       1150 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 361

Qy       1210 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 421

Qy       1270 LRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||
Db        422 LRRISENINKLRLGCNWQFIPVDEGWTE 449
```

RESULT 3
AAB04167
ID AAB04167 standard; protein; 449 AA.
XX
AC AAB04167;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype G).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54491.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 10b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 449 AA;

```
  Query Match             34.5%;  Score 448;  DB 4;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        850 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 61

Qy        910 NSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGW 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 NSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGW 121

Qy        970 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 181

Qy       1030 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 241

Qy       1090 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 301

Qy       1150 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 361

Qy       1210 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 421

Qy       1270 LRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||
Db        422 LRRISENINKLRLGCNWQFIPVDEGWTE 449
```

RESULT 4
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .21

```
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30596.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC   sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC   express BotG soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match             34.5%;  Score 447;  DB 2;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Qy        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146

Qy        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206

Qy       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266

Qy       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326

Qy       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386

Qy       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446

Qy       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||
Db        447 RRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 5
ABG69086
ID   ABG69086 standard; protein; 444 AA.
XX
AC   ABG69086;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #20.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98556.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 160-161; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 444 AA;

  Query Match             34.0%;  Score 441;  DB 5;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  441;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
```

```
Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMY 441
              |||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMY 441
```

```
RESULT 6
AAB04102
ID   AAB04102 standard; protein; 410 AA.
XX
AC   AAB04102;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain N-terminal sequence (serotype G).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54498.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 6; Fig 17b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 410 AA;

  Query Match             31.5%;  Score 408;  DB 4;  Length 410;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        442 KNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDL 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 KNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDL 62

Qy        502 SSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLT 561
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 SSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLT 122

Qy        562 NSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSD 621
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 NSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSD 182

Qy        622 VSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKG 681
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 VSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKG 242

Qy        682 HIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIED 741
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 HIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIED 302

Qy        742 QYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLK 801
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 QYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLK 362

Qy        802 DFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYT 849
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 DFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYT 410
```

```
RESULT 7
ABG69073
ID   ABG69073 standard; protein; 439 AA.
XX
AC   ABG69073;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #7.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
```

PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98543.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 129-130; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 439 AA;

```
  Query Match             20.4%;  Score 265;  DB 5;  Length 439;
  Best Local Similarity   99.7%;  Pred. No. 1.1e-262;
  Matches  385;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy         20 IMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFSKDVYEYYDPTYL 79
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 IMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFSKDVYEYYDPTYL 78

Qy         80 KTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPDKFAANVANVSIN 139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 KTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPDKFAANVANVSIN 138

Qy        140 KKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGFGARMMIRFCPSC 199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 KKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGFGARMMIRFCPSC 198

Qy        200 LNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNLPITPNTKEFFMQ 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 LNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNLPITPNTKEFFMQ 258

Qy        260 HSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLY 319
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        259 HSDPVQAEELYTFGGHDPSV-SPSTDMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLY 317

Qy        320 KQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPP 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 KQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPP 377

Qy        380 IKTEKLLDNTIYTQNEGFNIASKNLK 405
              |||||||||||||||||||||||||
Db        378 IKTEKLLDNTIYTQNEGFNIASKNLK 403
```

RESULT 8
ABG69068
ID ABG69068 standard; protein; 441 AA.
XX
AC ABG69068;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #2.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98538.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 119-120; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 441 AA;

```
  Query Match              1.2%;  Score 16;  DB 5;  Length 441;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243
```

RESULT 9
ABG69076
ID ABG69076 standard; protein; 441 AA.
XX
AC ABG69076;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #10.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;

KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98546.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 33; Page 135-136; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 441 AA;

Query Match 1.2%; Score 16; DB 5; Length 441;
Best Local Similarity 100.0%; Pred. No. 6.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 10
ABB80654
ID ABB80654 standard; peptide; 441 AA.
XX
AC ABB80654;
XX
DT 15-JUL-2002 (first entry)
XX
DE Botulinum toxin type B Danish I light chain.
XX
KW Neurotoxin; biological persistence; dysphonia; strabismus; muscle spasm;
KW dystonia; pain; blepharospasm; hemifacial spasm; excessive salivation;
KW eyelid disorder; cerebral palsy; focal spasticity; spasmodic colitis;
KW neurogenic bladder; anismus; limb spasticity; tic; tremor; bruxism;
KW anal fissure; achalasia; dysphagia; lacrimation; hyperhydrosis; headache;
KW excessive gastrointestinal secretion; botulinum toxin type B; Danish I;
KW light chain.
XX
OS Clostridium botulinum.
XX
PN WO200208268-A2.
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US023122.
XX
PR 21-JUL-2000; 2000US-00620840.
XX
PA (ALLR) ALLERGAN SALES INC.
XX
PI Steward LE, Fernandez-Salas E, Herrington TM, Aoki KR;
XX
DR WPI; 2002-241566/29.
XX
PT Novel modified neurotoxin comprising structural modification which alters
PT the biological persistence and/or biological activity of a neurotoxin,
PT useful for treating neuromuscular or autonomic disorder, or pain.
XX
PS Disclosure; Fig 8; 102pp; English.
XX
CC The sequence represents the botulinum toxin type B Danish I light chain.
CC The invention relates to a novel modified neurotoxin including a
CC structural modification, where the structural modification is effective
CC to alter the biological persistence, or biological activity. The modified
CC neurotoxin is useful for treating spasmodic dysphonia, laryngeal
CC dystonia, oromandibular dysphonia, lingual dystonia, cervical dystonia,
CC focal hand dystonia, blepharospasm, strabismus, hemifacial spasm, eyelid
CC disorder, cerebral palsy, focal spasticity, spasmodic colitis, neurogenic
CC bladder, anismus, limb spasticity, tics, tremors, bruxism, anal fissure,
CC achalasia, dysphagia, lacrimation, hyperhydrosis, excessive salivation,
CC excessive gastrointestinal secretions, pain from muscle spasms, headache
CC pain, brow furrows or skin wrinkles
XX
SQ Sequence 441 AA;

Query Match 1.2%; Score 16; DB 5; Length 441;
Best Local Similarity 100.0%; Pred. No. 6.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 11
ABU61545
ID ABU61545 standard; peptide; 441 AA.
XX
AC ABU61545;
XX
DT 11-AUG-2003 (first entry)
XX
DE Botulinum type toxin B Danish I light chain.
XX
KW Neurotoxin; analgesic; antiasthmatic; antiinflammatory; cerebral palsy;
KW biological persistence; biological persistence enhancing component;
KW leucine-based motif; neuromuscular disorder; autonomic disorder; pain;
KW spasmodic dysphonia; laryngeal dystonia; oromandibular dysphonia;
KW lingual dystonia; cervical dystonia; focal hand dystonia; blepharospasm;
KW strabismus; hemifacial spasm; eyelid disorder; focal spasticity;
KW spasmodic colitis; neurogenic bladder; anismus; limb spasticity; tics;
KW tremors; bruxism; anal fissure; achalasia; dysphagia; lacrimation;
KW hyperhydrosis; salivation; gastrointestinal secretion; muscle spasm;
KW headache pain; brow furrow; skin wrinkle; spinal curvature;
KW inflammatory pain; autonomic nervous system disorder; asthma; spasm;
KW respiratory malfunctioning; chronic obstructive pulmonary disease;
KW botulinum type toxin B Danish I light chain.
XX
OS Clostridium botulinum.
XX
PN US2003027752-A1.
XX
PD 06-FEB-2003.

XX
PF 20-JUL-2001; 2001US-00910346.
XX
PR 21-JUL-2000; 2000US-00620840.
XX
PA (ALLR) ALLERGAN SALES INC.
XX
PI Steward LE, Fernandez-Salas E, Herrington TM, Aoki KR;
XX
DR WPI; 2003-466155/44.
XX
PT Novel modified neurotoxin with a structural modification that alters
PT biological persistence or activity of the modified neurotoxin relative to
PT the unmodified neurotoxin, for treating tremors, bruxism and dysphagia.
XX
PS Disclosure; Fig 8; 33pp; English.
XX
CC The invention describes a modified neurotoxin (I) comprising a structural
CC modification that alters the biological persistence or biological
CC activity of the modified neurotoxin relative to an identical neurotoxin
CC without the structural modification (the modified neurotoxin is
CC structurally different from a naturally existing neurotoxin). (I) which
CC has altered biological persistence, is useful for treating a condition in
CC a mammal, where the neurotoxin does not comprise a leucine-based motif,
CC and the structural modification includes a biological persistence
CC enhancing component which comprises a leucine-based motif, tyrosine-based
CC motif or an amino acid derivative. (I) is useful for treating a condition
CC such as neuromuscular disorder, autonomic disorder or pain, or spasmodic
CC dysphonia, laryngeal dystonia, oromandibular dysphonia, lingual dystonia,
CC cervical dystonia, focal hand dystonia, blepharospasm, strabismus,
CC hemifacial spasm, eyelid disorder, cerebral palsy, focal spasticity,
CC spasmodic colitis, neurogenic bladder, anismus, limb spasticity, tics,
CC tremors, bruxism, anal fissure, achalasia, dysphagia, lacrimation,
CC hyperhydrosis, excessive salivation, excessive gastrointestinal
CC secretions, pain from muscle spasms, headache pain, brow furrows and skin
CC wrinkles. (I) is also useful for treating spinal curvature, various forms
CC of inflammatory pains, autonomic nervous system disorders, e,g., an
CC respiratory malfunctioning such as chronic obstructive pulmonary disease,
CC and asthma; pain not associated with a muscular disorder, such as spasm.
CC A unit amount of modified neurotoxin having altered biological activity
CC than natural neurotoxin, is more efficient to reduce exocytosis from a
CC cell than is a unit amount of naturally existing neurotoxin. This is the
CC amino acid sequence of a botulinum type toxin B Danish I light chain used
CC to determine a consensus sequence for Botulinum toxin
XX
SQ Sequence 441 AA;

Query Match 1.2%; Score 16; DB 6; Length 441;
Best Local Similarity 100.0%; Pred. No. 6.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 12
AAW56014
ID AAW56014 standard; protein; 548 AA.
XX
AC AAW56014;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH107/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26286.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS Example 2; Page 103-104; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 548 AA;

Query Match 1.2%; Score 16; DB 2; Length 548;
Best Local Similarity 100.0%; Pred. No. 8.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 13
ABG69077
ID ABG69077 standard; protein; 852 AA.
XX
AC ABG69077;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #11.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX

PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98547.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 138-139; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 852 AA;

Query Match 1.2%; Score 16; DB 5; Length 852;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 227 LMHELIHVLHGLYGIK 242

RESULT 14
AAW56018
ID AAW56018 standard; protein; 858 AA.
XX
AC AAW56018;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH417/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26290.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS Example 2; Page 98-100; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 858 AA;

Query Match 1.2%; Score 16; DB 2; Length 858;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 15
AAW56017
ID AAW56017 standard; protein; 1169 AA.
XX
AC AAW56017;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH728/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26289.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.

XX
PS Example 2; Page 91-94; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 1169 AA;

Query Match 1.2%; Score 16; DB 2; Length 1169;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

Search completed: September 10, 2004, 05:06:04
Job time : 77.9935 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 21.265 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 1297
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16	1.2	548	4	US-09-255-829-24	Sequence 24, Appl
2	16	1.2	858	4	US-09-255-829-22	Sequence 22, Appl
3	16	1.2	858	4	US-09-255-829-29	Sequence 29, Appl
4	16	1.2	1169	4	US-09-255-829-20	Sequence 20, Appl
5	14	1.1	1315	4	US-08-913-880C-1	Sequence 1, Appli
6	12	0.9	31	5	PCT-US93-11703-61	Sequence 61, Appl
7	11	0.8	207	3	US-08-446-114A-21	Sequence 21, Appl
8	11	0.8	452	1	US-07-618-312A-2	Sequence 2, Appli
9	11	0.8	452	1	US-07-618-312A-4	Sequence 4, Appli
10	11	0.8	452	1	US-08-110-786A-8	Sequence 8, Appli
11	11	0.8	452	1	US-08-280-228-2	Sequence 2, Appli
12	11	0.8	452	1	US-08-280-228-4	Sequence 4, Appli
13	11	0.8	618	1	US-08-668-381A-5	Sequence 5, Appli
14	11	0.8	853	4	US-08-913-880C-17	Sequence 17, Appl
15	11	0.8	858	4	US-08-913-880C-16	Sequence 16, Appl
16	11	0.8	860	4	US-08-913-880C-15	Sequence 15, Appl
17	11	0.8	862	4	US-08-913-880C-14	Sequence 14, Appl
18	11	0.8	865	4	US-08-913-880C-13	Sequence 13, Appl
19	11	0.8	866	4	US-08-913-880C-12	Sequence 12, Appl
20	11	0.8	871	4	US-09-255-829-2	Sequence 2, Appli
21	11	0.8	871	4	US-09-255-829-8	Sequence 8, Appli
22	11	0.8	871	4	US-09-255-829-26	Sequence 26, Appl
23	11	0.8	873	4	US-09-255-829-6	Sequence 6, Appli
24	11	0.8	874	4	US-08-913-880C-11	Sequence 11, Appl
25	11	0.8	875	4	US-08-913-880C-10	Sequence 10, Appl
26	11	0.8	875	4	US-09-255-829-10	Sequence 10, Appl
27	11	0.8	878	4	US-09-255-829-12	Sequence 12, Appl
28	11	0.8	894	4	US-09-255-829-4	Sequence 4, Appli
29	11	0.8	907	4	US-09-255-829-16	Sequence 16, Appl
30	11	0.8	953	4	US-09-255-829-14	Sequence 14, Appl
31	11	0.8	1013	4	US-09-255-829-18	Sequence 18, Appl
32	11	0.8	1296	1	US-08-480-604A-28	Sequence 28, Appl
33	11	0.8	1296	2	US-08-405-496A-28	Sequence 28, Appl
34	11	0.8	1296	3	US-08-915-136-28	Sequence 28, Appl
35	11	0.8	1296	4	US-09-084-517-28	Sequence 28, Appl
36	8	0.6	16	1	US-08-446-692-44	Sequence 44, Appl
37	8	0.6	16	2	US-08-488-351A-44	Sequence 44, Appl
38	8	0.6	16	5	PCT-US95-13841-16	Sequence 16, Appl
39	8	0.6	26	4	US-09-692-170C-19	Sequence 19, Appl
40	8	0.6	27	4	US-09-983-019-14	Sequence 14, Appl
41	8	0.6	28	4	US-09-983-019-11	Sequence 11, Appl
42	8	0.6	28	4	US-09-983-019-12	Sequence 12, Appl
43	8	0.6	28	4	US-09-983-019-13	Sequence 13, Appl
44	8	0.6	30	1	US-08-446-692-23	Sequence 23, Appl
45	8	0.6	30	2	US-08-488-351A-23	Sequence 23, Appl

ALIGNMENTS

```
RESULT 1
US-09-255-829-24
; Sequence 24, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-24

  Query Match             1.2%;  Score 16;  DB 4;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 6e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 2
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-22

  Query Match             1.2%;  Score 16;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 3
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             1.2%;  Score 16;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 4
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-20

  Query Match             1.2%;  Score 16;  DB 4;  Length 1169;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 5
US-08-913-880C-1
; Sequence 1, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
;  APPLICANT: MATSUDA, Morihiro
;  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
;  TITLE OF INVENTION:  VACCINE
;  FILE REFERENCE: 216-380P
;  CURRENT APPLICATION NUMBER: US/08/913,880C
;  CURRENT FILING DATE:  1997-11-24
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
;   LENGTH: 1315
;   TYPE: PRT
;   ORGANISM: Clostridium tetani
US-08-913-880C-1

  Query Match             1.1%;  Score 14;  DB 4;  Length 1315;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYG 241
              ||||||||||||||
Db        231 LMHELIHVLHGLYG 244


RESULT 6
PCT-US93-11703-61
; Sequence 61, Application PC/TUS9311703
;  GENERAL INFORMATION:
;    APPLICANT:  Chiron Mimotopes Pty. Ltd.
;    TITLE OF INVENTION:  T-Cell Epitopes
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Grant D. Green
;      STREET:  4560 Horton St.
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11703
;      FILING DATE:  28-DEC-1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/984,852
;      FILING DATE:  02-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Grant D.
;      REGISTRATION NUMBER:  31,259
;      REFERENCE/DOCKET NUMBER:  0222.101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2706
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-11703-61

  Query Match             0.9%;  Score 12;  DB 5;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 0.00044;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        156 TNLIIFGPGPVL 167
              ||||||||||||
Db         15 TNLIIFGPGPVL 26


RESULT 7
US-08-446-114A-21
; Sequence 21, Application US/08446114A
; Patent No. 6287566
;  GENERAL INFORMATION:
;    APPLICANT:  Dertzbaugh, Mark T.
;    TITLE OF INVENTION:  PROTECTIVE PEPTIDES OF NEUROTOXIN OF C.
;    TITLE OF INVENTION:  BOTULINUM
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hendricks and Associates
;      STREET:  9669 A Main Street
;      CITY:  Fairfax
;      STATE:  VA
;      COUNTRY:  US
;      ZIP:  22031
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,114A
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hendricks, Glenna
;      REGISTRATION NUMBER:  32,535
;      REFERENCE/DOCKET NUMBER:  der08/446,114
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-425-4250
;      TELEFAX:  703-425-2767
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  207 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  C. botulinum
```

```
US-08-446-114A-21

  Query Match             0.8%;  Score 11;  DB 3;  Length 207;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        624 IIIPYIGPALN 634
              ||||||||||
Db        178 IIIPYIGPALN 188


RESULT 8
US-07-618-312A-2
; Sequence 2, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             0.8%;  Score 11;  DB 1;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||
Db        253 DFWGNPLRYDT 263


RESULT 9
US-07-618-312A-4
; Sequence 4, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-4

  Query Match             0.8%;  Score 11;  DB 1;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||
Db        253 DFWGNPLRYDT 263


RESULT 10
US-08-110-786A-8
; Sequence 8, Application US/08110786A
; Patent No. 5443966
;  GENERAL INFORMATION:
;    APPLICANT:  FAIRWEATHER, Neil Fraser
;    APPLICANT:  MAKOFF, Andrew Joseph
;    TITLE OF INVENTION:  Expression of tetanus toxin fragment C
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye P.C.
;      STREET:  1100 No. 5443966th Glebe Road
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,786A
;      FILING DATE:  23-AUG-1993 1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/777,337
;      FILING DATE:  29-NOV-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB90/00943
;      FILING DATE:  20-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8914122.0
;      FILING DATE:  20 June 1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary J. Wilson
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  117-134
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-110-786A-8

  Query Match             0.8%;  Score 11;  DB 1;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||| 
Db        253 DFWGNPLRYDT 263


RESULT 11
US-08-280-228-2
; Sequence 2, Application US/08280228
; Patent No. 5571694
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. 5571694th Glebe Road
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,228
;      FILING DATE:  25-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/618,312
;      FILING DATE:  27-NOV-1990
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mary J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  117-163
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-280-228-2

  Query Match             0.8%;  Score 11;  DB 1;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||| 
Db        253 DFWGNPLRYDT 263


RESULT 12
US-08-280-228-4
; Sequence 4, Application US/08280228
; Patent No. 5571694
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. 5571694th Glebe Road
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,228
;      FILING DATE:  25-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/618,312
;      FILING DATE:  27-NOV-1990
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mary J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  117-163
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
```

```
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-280-228-4

  Query Match             0.8%;  Score 11;  DB 1;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||| 
Db        253 DFWGNPLRYDT 263


RESULT 13
US-08-668-381A-5
; Sequence 5, Application US/08668381A
; Patent No. 5780024
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Robert H.
;    APPLICANT:  Fishman, Paul S.
;    APPLICANT:  Francis, Jonathan W.
;    APPLICANT:  Hosler, Betsy A.
;    TITLE OF INVENTION:  SUPEROXIDE DISMUTASE/TETANUS TOXIN
;    TITLE OF INVENTION:  FRAGMENT C HYBRID PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,381A
;      FILING DATE:  21-JUN-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/000,473
;      FILING DATE:  23-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,164
;      REFERENCE/DOCKET NUMBER:  00786/269001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  618 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-668-381A-5

  Query Match             0.8%;  Score 11;  DB 1;  Length 618;
  Best Local Similarity   100.0%;  Pred. No. 0.051;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||| 
Db        419 DFWGNPLRYDT 429


RESULT 14
US-08-913-880C-17
; Sequence 17, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
;  APPLICANT: MATSUDA, Morihiro
;  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
;  TITLE OF INVENTION:  VACCINE
;  FILE REFERENCE: 216-380P
;  CURRENT APPLICATION NUMBER: US/08/913,880C
;  CURRENT FILING DATE:  1997-11-24
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 17
;   LENGTH: 853
;   TYPE: PRT
;   ORGANISM: Clostridium tetani
;   FEATURE:
;   OTHER INFORMATION: Amino Acids 463 to 1315 of SEQ ID NO: 1
US-08-913-880C-17

  Query Match             0.8%;  Score 11;  DB 4;  Length 853;
  Best Local Similarity   100.0%;  Pred. No. 0.067;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||| 
Db        654 DFWGNPLRYDT 664


RESULT 15
US-08-913-880C-16
; Sequence 16, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
;  APPLICANT: MATSUDA, Morihiro
;  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
;  TITLE OF INVENTION:  VACCINE
;  FILE REFERENCE: 216-380P
;  CURRENT APPLICATION NUMBER: US/08/913,880C
;  CURRENT FILING DATE:  1997-11-24
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 16
;   LENGTH: 858
;   TYPE: PRT
;   ORGANISM: Clostridium tetani
;   FEATURE:
;   OTHER INFORMATION: Amino Acids 458 to 1315 of SEQ ID NO: 1
US-08-913-880C-16

  Query Match             0.8%;  Score 11;  DB 4;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 0.067;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1094 DFWGNPLRYDT 1104
              ||||||||||| 
Db        659 DFWGNPLRYDT 669


Search completed: September 10, 2004, 05:17:10
Job time : 22.265 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 65.6083 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 1297
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1297	100.0	1297	12	US-10-272-898-77	Sequence 77, Appl
2	1297	100.0	1297	15	US-10-354-774-77	Sequence 77, Appl
3	1297	100.0	1297	15	US-10-271-012-77	Sequence 77, Appl
4	1297	100.0	1297	16	US-10-729-122-77	Sequence 77, Appl
5	1297	100.0	1297	16	US-10-729-039-77	Sequence 77, Appl
6	1297	100.0	2017	15	US-10-452-024-146	Sequence 146, App
7	1290	99.5	1297	15	US-10-452-024-7	Sequence 7, Appli
8	1290	99.5	1297	15	US-10-452-024-150	Sequence 150, App
9	1290	99.5	1297	15	US-10-205-516-14	Sequence 14, Appl
10	990	76.3	1307	15	US-10-205-516-28	Sequence 28, Appl
11	728	56.1	848	13	US-10-011-588-45	Sequence 45, Appl
12	448	34.5	449	10	US-09-910-186A-18	Sequence 18, Appl
13	447	34.5	473	12	US-10-272-898-79	Sequence 79, Appl
14	447	34.5	473	15	US-10-354-774-79	Sequence 79, Appl
15	447	34.5	473	15	US-10-271-012-79	Sequence 79, Appl
16	447	34.5	473	16	US-10-729-122-79	Sequence 79, Appl
17	447	34.5	473	16	US-10-729-039-79	Sequence 79, Appl
18	441	34.0	444	13	US-10-011-588-43	Sequence 43, Appl
19	408	31.5	410	10	US-09-910-186A-32	Sequence 32, Appl
20	265	20.4	439	13	US-10-011-588-17	Sequence 17, Appl
21	16	1.2	211	15	US-10-452-024-163	Sequence 163, App
22	16	1.2	424	15	US-10-452-024-138	Sequence 138, App
23	16	1.2	425	15	US-10-452-024-137	Sequence 137, App
24	16	1.2	440	15	US-10-163-106B-20	Sequence 20, Appl
25	16	1.2	441	10	US-09-910-346C-20	Sequence 20, Appl
26	16	1.2	441	13	US-10-011-588-7	Sequence 7, Appli
27	16	1.2	441	13	US-10-011-588-23	Sequence 23, Appl
28	16	1.2	451	15	US-10-452-024-133	Sequence 133, App
29	16	1.2	458	14	US-10-241-596-114	Sequence 114, App
30	16	1.2	548	14	US-10-241-596-24	Sequence 24, Appl
31	16	1.2	852	13	US-10-011-588-25	Sequence 25, Appl
32	16	1.2	858	14	US-10-241-596-22	Sequence 22, Appl
33	16	1.2	860	14	US-10-241-596-175	Sequence 175, App
34	16	1.2	862	14	US-10-241-596-94	Sequence 94, Appl
35	16	1.2	862	14	US-10-241-596-171	Sequence 171, App
36	16	1.2	862	14	US-10-241-596-173	Sequence 173, App
37	16	1.2	864	14	US-10-241-596-102	Sequence 102, App
38	16	1.2	865	14	US-10-241-596-100	Sequence 100, App
39	16	1.2	866	14	US-10-241-596-88	Sequence 88, Appl
40	16	1.2	866	14	US-10-241-596-104	Sequence 104, App
41	16	1.2	867	14	US-10-241-596-80	Sequence 80, Appl
42	16	1.2	867	14	US-10-241-596-96	Sequence 96, Appl
43	16	1.2	867	14	US-10-241-596-98	Sequence 98, Appl
44	16	1.2	870	14	US-10-241-596-92	Sequence 92, Appl
45	16	1.2	871	14	US-10-241-596-84	Sequence 84, Appl

ALIGNMENTS

```
RESULT 1
US-10-272-898-77
; Sequence 77, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 77:
;       SEQUENCE CHARACTERISTICS:
```

```
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-272-898-77

  Query Match           100.0%;  Score 1297;  DB 12;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 2
US-10-354-774-77
; Sequence 77, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-354-774-77

  Query Match           100.0%;  Score 1297;  DB 15;  Length 1297;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLPEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 3
US-10-271-012-77
; Sequence 77, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-271-012-77

  Query Match           100.0%;  Score 1297;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
```

```
Db        61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Qy       121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db       121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Qy       181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db       181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Qy       241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db       241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Qy       301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
Db       301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
Qy       361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
Db       361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
Qy       421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
Db       421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
Qy       481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
Db       481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
Qy       541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
Db       541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
Qy       601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
Db       601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
Qy       661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
Db       661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
Qy       721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
Db       721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
Qy       781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
Db       781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
Qy       841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
Db       841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
Qy       901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
Db       901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
Qy       961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
Db       961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
Qy      1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
Db      1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
Qy      1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
Db      1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
Qy      1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
Db      1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
Qy      1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
Db      1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
Qy      1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
Db      1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

```
RESULT 4
US-10-729-122-77
; Sequence 77, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-122-77

  Query Match           100.0%;  Score 1297;  DB 16;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
                 |||||||||||||||||||||||||||||||||||||
```

```
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 5
US-10-729-039-77
; Sequence 77, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-039-77

  Query Match           100.0%;  Score 1297;  DB 16;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
```

```
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 6
US-10-452-024-146
; Sequence 146, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 146
; LENGTH: 2017
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-146

Query Match 100.0%; Score 1297; DB 15; Length 2017;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db        721 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 780

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db        781 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 840

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Db        841 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 900

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Db        901 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 960

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Db        961 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 1020

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Db       1021 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 1080

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Db       1081 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 1140

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Db       1141 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 1200

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Db       1201 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 1260

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Db       1261 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 1320

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Db       1321 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 1380

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Db       1381 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 1440

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Db       1441 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 1500
```

```
Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 1560

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 1620

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 1680

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1740

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1800

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1860

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1920

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1980

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1981 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 2017


RESULT 7
US-10-452-024-7
; Sequence 7, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-7

  Query Match             99.5%;  Score 1290;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67

Qy         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127

Qy        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187

Qy        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247

Qy        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307

Qy        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367

Qy        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427

Qy        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487

Qy        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547

Qy        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607

Qy        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667

Qy        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727

Qy        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787

Qy        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847

Qy        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907

Qy        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967

Qy        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027

Qy       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087

Qy       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147

Qy       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 1207
```

```
Qy       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267

Qy       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
             ||||||||||||||||||||||||||||||
Db       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 8
US-10-452-024-150
; Sequence 150, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 150
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-150

  Query Match             99.5%;  Score 1290;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67

Qy         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127

Qy        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187

Qy        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247

Qy        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307

Qy        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367

Qy        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427

Qy        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487

Qy        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547

Qy        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607

Qy        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667

Qy        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727

Qy        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787

Qy        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847

Qy        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907

Qy        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967

Qy        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027

Qy       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087

Qy       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147

Qy       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 1207

Qy       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267

Qy       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||
Db       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 9
US-10-205-516-14
; Sequence 14, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
```

```
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7)
;   OTHER INFORMATION: mixed sequence
US-10-205-516-14

  Query Match            99.5%;  Score 1290;  DB 15;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67

Qy         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127

Qy        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187

Qy        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247

Qy        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307

Qy        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367

Qy        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427

Qy        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487

Qy        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547

Qy        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607

Qy        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667

Qy        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727

Qy        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787

Qy        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847

Qy        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907

Qy        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967

Qy        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027

Qy       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087

Qy       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147

Qy       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 1207

Qy       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267

Qy       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||
Db       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 10
US-10-205-516-28
; Sequence 28, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
;   LENGTH: 1307
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7)
;   OTHER INFORMATION: mixed sequence
US-10-205-516-28

  Query Match            76.3%;  Score 990;  DB 15;  Length 1307;
  Best Local Similarity  99.7%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    4;  Gaps    1;

Qy          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67

Qy         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127

Qy        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187

Qy        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247

Qy        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307
```

```
Qy        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367

Qy        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427

Qy        428 LVIYRIAMCKPVMYK----NTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQN 483
              |||||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db        428 LVIYRIAMCKPVMYKIEGRNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQN 487

Qy        484 NTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLF 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 NTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLF 547

Qy        544 EYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVI 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 EYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVI 607

Qy        604 DDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPEL 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 DDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPEL 667

Qy        664 IVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKER 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 IVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKER 727

Qy        724 MYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQC 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 MYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQC 787

Qy        784 SISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPF 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 SISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPF 847

Qy        844 DLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGN 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 DLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGN 907

Qy        904 GQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCI 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 GQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCI 967

Qy        964 KNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNAN 1023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 KNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNAN 1027

Qy       1024 IYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLY 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 IYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLY 1087

Qy       1084 WIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQ 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 WIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQ 1147

Qy       1144 NLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFL 1203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 NLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFL 1207

Qy       1204 APINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYF 1263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 APINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYF 1267

Qy       1264 CISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||||
Db       1268 CISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1301
```

RESULT 11
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1

```
; GENERAL INFORMATION:
;  APPLICANT: Smith, Leonard
;  APPLICANT:  Jensen, Melody
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 848
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45
```

Query Match 56.1%; Score 728; DB 13; Length 848;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 848; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIEN-QLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 599

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        600 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 659

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 719

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 779

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 839

Qy        841 IPFDLSLYT 849
              |||||||||
Db        840 IPFDLSLYT 848


RESULT 12
US-09-910-186A-18
; Sequence 18, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 449
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-09-910-186A-18

  Query Match           34.5%;  Score 448;  DB 10;  Length 449;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        850 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 61

Qy        910 NSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGW 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 NSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGW 121

Qy        970 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 181

Qy       1030 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 241

Qy       1090 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 301

Qy       1150 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 361

Qy       1210 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 421

Qy       1270 LRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||
Db        422 LRRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 13
US-10-272-898-79
; Sequence 79, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 79:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 473 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-272-898-79

  Query Match           34.5%;  Score 447;  DB 12;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Qy        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146

Qy        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206

Qy       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266

Qy       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326

Qy       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386

Qy       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446

Qy       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||
Db        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
```

```
RESULT 14
US-10-354-774-79
; Sequence 79, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-354-774-79

  Query Match           34.5%;  Score 447;  DB 15;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Qy        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146

Qy        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206

Qy       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266

Qy       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326

Qy       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386

Qy       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446

Qy       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||
Db        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
```

```
RESULT 15
US-10-271-012-79
; Sequence 79, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;
```

```
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 473 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-271-012-79

  Query Match            34.5%;  Score 447;  DB 15;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Qy        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146

Qy        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206

Qy       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266

Qy       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326

Qy       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386

Qy       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446

Qy       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||
Db        447 RRISENINKLRLGCNWQFIPVDEGWTE 473


Search completed: September 10, 2004, 05:53:03
Job time : 69.6083 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 17.9681 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 1297
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1297	100.0	1297	2	S39791	neurotoxin - Clost
2	16	1.2	1291	1	A48940	bontoxilysin (EC 3
3	16	1.2	1291	2	I40631	non-proteolytic bo
4	14	1.1	1315	1	BTCLTN	tentoxilysin (EC 3
5	12	0.9	1251	2	JH0256	botulinum neurotox
6	12	0.9	1252	2	S21178	botulinum neurotox
7	12	0.9	1268	2	S33411	botulinum neurotox
8	11	0.8	1296	1	BTCLAB	bontoxilysin (EC 3
9	10	0.8	1276	2	S11455	botulinum neurotox
10	10	0.8	1285	2	S70582	botulinum neurotox
11	10	0.8	1291	2	S46431	botulinum neurotox
12	10	0.8	1291	2	A49777	botulinum neurotox
13	9	0.7	1148	2	A71446	hypothetical prote
14	8	0.6	229	2	D38941	cytochrome-c oxida
15	8	0.6	286	2	T39404	hypothetical prote
16	8	0.6	638	2	AF3463	flagellar protein
17	8	0.6	738	2	T19938	hypothetical prote
18	8	0.6	835	1	A49891	outer membrane ush
19	8	0.6	980	1	S45444	BEM1 protein-bindi
20	8	0.6	1296	2	I40645	botulinum neurotox
21	8	0.6	1386	2	T00257	hypothetical prote
22	7	0.5	95	2	F72864	AcOrf-117 protein
23	7	0.5	95	2	T41855	AcMNPV orf117 - Bo
24	7	0.5	98	2	I40388	late competence ge
25	7	0.5	125	2	C70162	ribosomal protein
26	7	0.5	125	2	S76368	hypothetical prote
27	7	0.5	135	2	S34546	H+-transporting tw
28	7	0.5	138	2	S71145	H+-transporting tw
29	7	0.5	139	2	A69052	riboflavin synthas
30	7	0.5	141	2	H64337	hypothetical prote
31	7	0.5	144	2	A61158	aminoglycoside N6'
32	7	0.5	155	2	C81150	hypothetical prote
33	7	0.5	159	2	S20143	hypothetical prote
34	7	0.5	160	2	G82414	creA protein VCA08
35	7	0.5	162	2	T14635	hypothetical prote
36	7	0.5	178	2	B81872	hypothetical prote
37	7	0.5	186	2	AI2405	hypothetical prote
38	7	0.5	189	2	AH1029	probable membrane
39	7	0.5	190	2	H69778	hypothetical prote
40	7	0.5	195	2	T46701	hypothetical prote
41	7	0.5	196	2	PD0004	self-incompatibili
42	7	0.5	200	2	D90191	conserved hypothet
43	7	0.5	203	2	G95309	conserved hypothet
44	7	0.5	205	2	H69854	hypothetical prote
45	7	0.5	214	2	E70439	ATP phosphoribosyl

ALIGNMENTS

RESULT 1
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 100.0%; Score 1297; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 2
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisio
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
```

F;231/Active site: Glu #status predicted

Query Match 1.2%; Score 16; DB 1; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 3
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 1.2%; Score 16; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYGIK 243
||||||||||||||||
Db 228 LMHELIHVLHGLYGIK 243

RESULT 4
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in H
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglic
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 1.1%; Score 14; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 LMHELIHVLHGLYG 241
||||||||||||||
Db 231 LMHELIHVLHGLYG 244

RESULT 5
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match 0.9%; Score 12; DB 2; Length 1251;
  Best Local Similarity 100.0%; Pred. No. 0.0028;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 DPALTLMHELIH 234
Db 205 DPALTLMHELIH 216

RESULT 6
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-11
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neur
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match 0.9%; Score 12; DB 2; Length 1252;
  Best Local Similarity 100.0%; Pred. No. 0.0028;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 DPALTLMHELIH 234
Db 205 DPALTLMHELIH 216

RESULT 7
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 0.9%; Score 12; DB 2; Length 1268;
  Best Local Similarity 100.0%; Pred. No. 0.0028;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 230 HELIHVLHGLYG 241
Db 219 HELIHVLHGLYG 230

RESULT 8
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999

C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, ar
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

  Query Match 0.8%; Score 11; DB 1; Length 1296;
  Best Local Similarity 100.0%; Pred. No. 0.032;
  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 624 IIIPYIGPALN 634
        ||||||||||
Db 632 IIIPYIGPALN 642

RESULT 9
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegl
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match 0.8%; Score 10; DB 2; Length 1276;
  Best Local Similarity 100.0%; Pred. No. 0.36;
  Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 625 IIPYIGPALN 634
        ||||||||||
Db 624 IIPYIGPALN 633

RESULT 10
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C a
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit ne
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic memb
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match 0.8%; Score 10; DB 2; Length 1285;
  Best Local Similarity 100.0%; Pred. No. 0.36;
  Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 625 IIPYIGPALN 634
        ||||||||||
Db 624 IIPYIGPALN 633

RESULT 11

S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 0.8%; Score 10; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 625 IIPYIGPALN 634
||||||||||
Db 628 IIPYIGPALN 637

RESULT 12
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 0.8%; Score 10; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 625 IIPYIGPALN 634
||||||||||
Db 628 IIPYIGPALN 637

RESULT 13
A71446
hypothetical protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C;Accession: A71446
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: A71446
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1148 <BEV>
A;Cross-references: GB:Z97343; NID:g2245073; PID:e327073; PID:g2245120
C;Genetics:
A;Map position: 4COP9-4G3845

Query Match 0.7%; Score 9; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 398 NIASKNLKT 406
|||||||||
Db 590 NIASKNLKT 598

RESULT 14
D38941
cytochrome-c oxidase (EC 1.9.3.1) chain II - dragonfly (Sympetrum striolatum) mitochondr
C;Species: mitochondrion Sympetrum striolatum (dragonfly)
C;Date: 18-Jul-1995 #sequence_revision 28-Jul-1995 #text_change 18-Aug-2003
C;Accession: D38941
R;Liu, H.; Beckenbach, A.T.
Mol. Phylogenet. Evol. 1, 41-52, 1992
A;Title: Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of in
A;Reference number: A45170; MUID:94115687; PMID:1342923
A;Accession: D38941
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <LIU>
A;Cross-references: GB:M83962; NID:g343437; PIDN:AAA32087.1; PID:g552893
C;Genetics:
A;Gene: COII
A;Genome: mitochondrion
A;Genetic code: SGC4
C;Superfamily: cytochrome-c oxidase chain II, mitochondrial type; cytochrome-c oxidase c
C;Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner
F;9-214/Domain: cytochrome-c oxidase chain II homology <CO2>
F;161,196,200,207/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F;196,198,200,204/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F;198/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 0.6%; Score 8; DB 2; Length 229;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 820 LYLLDEVN 827

```
                ||||||||
Db         84 LYLLDEVN 91


RESULT 15
T39404
hypothetical protein SPBC13G1.03c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39404
R;Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z21852
A;Accession: T39404
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-286 <LYN>
A;Cross-references: EMBL:AL022600; PIDN:CAA18656.1; GSPDB:GN00067; SPDB:SPBC13G1.03c
A;Experimental source: strain 972h-; cosmid c13G1
C;Genetics:
A;Gene: SPDB:SPBC13G1.03c
A;Map position: 2
A;Introns: 91/3

  Query Match             0.6%;  Score 8;  DB 2;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        755 NIDFNDID 762
                ||||||||
Db        268 NIDFNDID 275


Search completed: September 10, 2004, 05:14:56
Job time : 20.0681 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 11.2095 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 1297
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	99.5	1296	1	BXG_CLOBO	Q60393 clostridium
2	16	1.2	1290	1	BXB_CLOBO	P10844 clostridium
3	14	1.1	1314	1	TETX_CLOTE	P04958 clostridium
4	12	0.9	1250	1	BXE_CLOBO	Q00496 clostridium
5	12	0.9	1250	1	BXE_CLOBU	P30995 clostridium
6	11	0.8	1295	1	BXA1_CLOBO	P10845 clostridium
7	10	0.8	1276	1	BXD_CLOBO	P19321 clostridium
8	10	0.8	1290	1	BXC1_CLOBO	P18640 clostridium
9	8	0.6	229	1	COX2_SYMST	P29880 sympetrum s
10	8	0.6	835	1	FASD_ECOLI	P46000 escherichia
11	8	0.6	980	1	BOB1_YEAST	P38041 saccharomyc
12	8	0.6	1295	1	BXA2_CLOBO	Q45894 clostridium
13	7	0.5	95	1	Y117_NPVAC	P41670 autographa
14	7	0.5	98	1	CMF2_BACSU	P39146 bacillus su
15	7	0.5	125	1	RS13_BORBU	O51453 borrelia bu
16	7	0.5	135	1	ATPE_EUGGR	P31477 euglena gra
17	7	0.5	138	1	ATPE_ANTFO	Q31793 anthoceros
18	7	0.5	139	1	RISB_METTH	O27443 methanobact
19	7	0.5	141	1	RISB_METJA	Q57751 methanococc
20	7	0.5	147	1	YN28_VIBPA	Q87mc8 vibrio para
21	7	0.5	149	1	YI47_VIBVU	Q8dbh0 vibrio vuln
22	7	0.5	157	1	RISB_SULTO	Q975m5 sulfolobus
23	7	0.5	167	1	CATR_ATRNU	P41210 atriplex nu
24	7	0.5	194	1	KITH_BACAA	Q81jx0 bacillus an
25	7	0.5	195	1	KITH_BACCR	Q814u0 bacillus ce
26	7	0.5	205	1	YKAA_BACSU	O34454 bacillus su
27	7	0.5	214	1	CYB_CERCE	P87419 cerastes ce
28	7	0.5	214	1	CYB_LACMU	P92853 lachesis mu
29	7	0.5	214	1	HIS1_AQUAE	O67543 aquifex aeo
30	7	0.5	218	1	RNS7_NICAL	Q40381 nicotiana a
31	7	0.5	226	1	YG64_METJA	Q59058 methanococc
32	7	0.5	231	1	RSUA_VIBCH	Q9krk5 vibrio chol
33	7	0.5	248	1	COX2_SCHPO	P21534 schizosacch
34	7	0.5	248	1	TPIS_MYCPU	Q98qa8 mycoplasma
35	7	0.5	259	1	UL31_HSV7J	P52361 human herpe
36	7	0.5	260	1	Y507_METKA	Q8ty00 methanopyru
37	7	0.5	270	1	Y304_MYCPN	P75355 mycoplasma
38	7	0.5	274	1	ATP6_BUCAI	P57118 buchnera ap
39	7	0.5	305	1	OTCC_THEAC	Q9hik9 thermoplasm
40	7	0.5	316	1	PGLB_CHICK	Q90944 gallus gall
41	7	0.5	320	1	PR7R_MMTVG	P10261 mouse mamma
42	7	0.5	324	1	PR7L_MMTVG	P03321 mouse mamma
43	7	0.5	334	1	KIME_PYRFU	Q8u0f3 pyrococcus
44	7	0.5	342	1	DHYS_PYRHO	O50105 pyrococcus
45	7	0.5	342	1	IDI2_RICCN	Q92hm7 rickettsia

ALIGNMENTS

RESULT 1
BXG_CLOBO
ID BXG_CLOBO STANDARD; PRT; 1296 AA.
AC Q60393;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE (Bontoxilysin G).
GN BOTG.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;
RX MEDLINE=94092745; PubMed=8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins.";
RL Biochim. Biophys. Acta 1216:487-491(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; X74162; CAA52275.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.

```
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match           99.5%;  Score 1290;  DB 1;  Length 1296;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7 FNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGVFS 66

Qy         68 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 KDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNASTPPD 126

Qy        128 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGF 186

Qy        188 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIKISNL 246

Qy        248 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 PITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVS 306

Qy        308 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 SAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIK 366

Qy        368 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEH 426

Qy        428 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 LVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIE 486

Qy        488 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 NNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLH 546

Qy        548 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 AQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFT 606

Qy        608 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 SESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPI 666

Qy        668 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 VGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNA 726

Qy        728 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 LNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISY 786

Qy        788 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 LMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSL 846

Qy        848 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 YTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFK 906

Qy        908 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 LNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDS 966

Qy        968 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 GWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIN 1026

Qy       1028 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQS 1086

Qy       1088 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 STNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYL 1146

Qy       1148 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPIN 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1147 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPIN 1206

Qy       1208 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1207 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 1266

Qy       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||
Db       1267 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1296


RESULT 2
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
```

```
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match              1.2%;  Score 16;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        227 LMHELIHVLHGLYGIK 242


RESULT 3
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
```

```
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
```

```
FT   STRAND      1127   1131
FT   HELIX       1132   1134
FT   TURN        1135   1136
FT   STRAND      1137   1141
FT   TURN        1144   1145
FT   STRAND      1148   1152
FT   STRAND      1155   1158
FT   TURN        1159   1162
FT   STRAND      1163   1166
FT   STRAND      1173   1178
FT   TURN        1184   1185
FT   STRAND      1188   1188
FT   STRAND      1190   1190

  Query Match             1.1%;  Score 14;  DB 1;  Length 1314;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYG 241
              ||||||||||||||
Db        230 LMHELIHVLHGLYG 243


RESULT 4
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
```

```
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             0.9%;  Score 12;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        223 DPALTLMHELIH 234
              ||||||||||||
Db        204 DPALTLMHELIH 215


RESULT 5
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             0.9%;  Score 12;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        223 DPALTLMHELIH 234
              ||||||||||||
Db        204 DPALTLMHELIH 215


RESULT 6
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
```

```
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
```

```
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             0.8%;  Score 11;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.029;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        624 IIIPYIGPALN 634
              ||||||||||| 
Db        631 IIIPYIGPALN 641


RESULT 7
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             0.8%;  Score 10;  DB 1;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        625 IIPYIGPALN 634
              ||||||||||
```

```
Db        624 IIPYIGPALN 633


RESULT 8
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             0.8%;  Score 10;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        625 IIPYIGPALN 634
              ||||||||||
Db        627 IIPYIGPALN 636


RESULT 9
COX2_SYMST
ID   COX2_SYMST     STANDARD;      PRT;   229 AA.
AC   P29880;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome c oxidase polypeptide II (EC 1.9.3.1).
GN   COII.
OS   Sympetrum striolatum (Common darter).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Palaeoptera; Odonata; Anisoptera; Libellulidae; Sympetrum.
OX   NCBI_TaxID=6969;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu H., Beckenbach A.T.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory
CC       chain that catalyzes the reduction of oxygen to water. Subunits 1-
CC       3 form the functional core of the enzyme complex. Subunit 2
CC       transfers the electrons from cytochrome c via its binuclear copper
CC       A center to the bimetallic center of the catalytic subunit 1.
CC   -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
CC       c + 2 H(2)O.
CC   -!- COFACTOR: Copper A.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC       inner membrane.
CC   -!- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M83962; AAA32087.1; -.
DR   PIR; D38941; D38941.
DR   InterPro; IPR001505; Copper_CuA.
DR   InterPro; IPR008972; Cupredoxin.
DR   InterPro; IPR002429; Cyt_c_ox_2.
DR   Pfam; PF00116; COX2; 1.
DR   Pfam; PF02790; COX2_TM; 1.
DR   PRINTS; PR01166; CYCOXIDASEII.
DR   ProDom; PD000131; Copper_CuA; 1.
DR   PROSITE; PS00078; COX2; 1.
KW   Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
KW   Electron transport; Respiratory chain.
FT   DOMAIN        1     26       MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT   TRANSMEM     27     48       POTENTIAL.
FT   DOMAIN       49     62       MITOCHONDRIAL MATRIX (POTENTIAL).
FT   TRANSMEM     63     82       POTENTIAL.
FT   DOMAIN       83    229       MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT   METAL       161    161       COPPER A (PROBABLE).
FT   METAL       196    196       COPPER A (PROBABLE).
FT   METAL       200    200       COPPER A (PROBABLE).
FT   METAL       204    204       COPPER A (PROBABLE).
SQ   SEQUENCE   229 AA;  26565 MW;  2581DEAFDB549473 CRC64;

  Query Match              0.6%;  Score 8;  DB 1;  Length 229;
  Best Local Similarity  100.0%;  Pred. No. 7.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        820 LYLLDEVN 827
              ||||||||
Db         84 LYLLDEVN 91


RESULT 10
FASD_ECOLI
ID   FASD_ECOLI     STANDARD;      PRT;   835 AA.
AC   P46000;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Outer membrane usher protein fasD precursor.
GN   FASD.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=987;
RX   MEDLINE=94148769; PubMed=7906265;
RA   Schifferli D.M., Alrutz M.A.;
RT   "Permissive linker insertion sites in the outer membrane protein of
RT   987P fimbriae of Escherichia coli.";
RL   J. Bacteriol. 176:1099-1110(1994).
CC   -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE 987P
CC       FIMBRIAE SUBUNITS ACROSS THE OUTER MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC   -!- SIMILARITY: Belongs to the fimbrial export usher family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L22659; AAA21827.1; -.
DR   EMBL; U50547; AAB02687.1; -.
DR   PIR; A49891; A49891.
DR   InterPro; IPR000015; Fimb_usher.
DR   Pfam; PF00577; Usher; 1.
DR   PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW   Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    835       OUTER MEMBRANE USHER PROTEIN FASD.
FT   DISULFID    810    834       POTENTIAL.
SQ   SEQUENCE   835 AA;  92354 MW;  2FECE6E2305274E1 CRC64;

  Query Match              0.6%;  Score 8;  DB 1;  Length 835;
  Best Local Similarity  100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        567 ALRNNNKV 574
              ||||||||
Db        549 ALRNNNKV 556


RESULT 11
BOB1_YEAST
ID   BOB1_YEAST     STANDARD;      PRT;   980 AA.
AC   P38041;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   BOB1 protein (BEM1-binding protein).
GN   BOB1 OR BOI1 OR YBL085W OR YBL0717.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96234675; PubMed=8666672;
RA   Bender L., Lo H.S., Lee H., Kokojan V., Peterson V., Bender A.;
RT   "Associations among PH and SH3 domain-containing proteins and
RT   Rho-type GTPases in Yeast.";
RL   J. Cell Biol. 133:879-894(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=96076635; PubMed=7502586;
RA   Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
RT   "Sequence analysis of a 78.6 kb segment of the left end of
RT   Saccharomyces cerevisiae chromosome II.";
RL   Yeast 11:1103-1112(1995).
CC   -!- FUNCTION: Binds to the BEM1 protein.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L31406; AAB08439.1; -.
DR   EMBL; X79489; CAA56021.1; -.
DR   EMBL; Z35846; CAA84906.1; -.
DR   PIR; S45444; S45444.
DR   GermOnline; 138520; -.
DR   SGD; S0000181; BOI1.
DR   GO; GO:0005935; C:bud neck; IDA.
DR   GO; GO:0005543; F:phospholipid binding; IDA.
DR   GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IGI.
DR   GO; GO:0007266; P:Rho protein signal transduction; IGI.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR001660; SAM.
```

```
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00536; SAM; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00454; SAM; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50105; SAM_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain.
FT   DOMAIN       13     77       SH3.
FT   DOMAIN      228    292       SAM.
FT   DOMAIN      776    895       PH.
SQ   SEQUENCE   980 AA;  109295 MW;  09F1DD1F9EF30F36 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 980;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        862 ISNISSNA 869
              ||||||||
Db        498 ISNISSNA 505


RESULT 12
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             0.6%;  Score 8;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        627 PYIGPALN 634
              ||||||||
Db        634 PYIGPALN 641


RESULT 13
Y117_NPVAC
ID   Y117_NPVAC     STANDARD;      PRT;    95 AA.
AC   P41670;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 11.0 kDa protein in HE65-PK2 intergenic region.
OS   Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=46015;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C6;
RX   MEDLINE=94303173; PubMed=8030224;
RA   Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT   "The complete DNA sequence of Autographa californica nuclear
RT   polyhedrosis virus.";
RL   Virology 202:586-605(1994).
CC   -!- SIMILARITY: TO CORRESPONDING ORF IN OPMNPV.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L22858; AAA66747.1; -.
DR   PIR; F72864; F72864.
KW   Hypothetical protein.
SQ   SEQUENCE   95 AA;  10992 MW;  68897B745ECE51C6 CRC64;

  Query Match             0.5%;  Score 7;  DB 1;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        689 NALKKRD 695
              |||||||
Db         12 NALKKRD 18


RESULT 14
CMF2_BACSU
ID   CMF2_BACSU     STANDARD;      PRT;    98 AA.
AC   P39146;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   ComF operon protein 2.
GN   COMFB OR COMF2 OR BSU35460.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=94018599; PubMed=8412657;
RA   Londono-Vallejo J.A., Dubnau D.;
RT   "comF, a Bacillus subtilis late competence locus, encodes a protein
RT   similar to ATP-dependent RNA/DNA helicases.";
RL   Mol. Microbiol. 9:119-131(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Soldo B., Lazarevic V., Mauel C., Karamata D.;
RT   "Sequence of the Bacillus subtilis 168 chromosomal region from 305
RT   to 307 degree.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z18629; CAA79227.1; -.
DR   EMBL; U56901; AAC44941.1; -.
DR   EMBL; Z99122; CAB15563.1; -.
DR   PIR; I40388; I40388.
DR   SubtiList; BG10396; comFB.
KW   Competence; Complete proteome.
SQ   SEQUENCE   98 AA;  10976 MW;  B365DB252B25D9F7 CRC64;

  Query Match             0.5%;  Score 7;  DB 1;  Length 98;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1191 LVNSKEI 1197
              |||||||
Db          2 LVNSKEI 8


RESULT 15
RS13_BORBU
ID   RS13_BORBU     STANDARD;      PRT;   125 AA.
AC   O51453;
DT   15-DEC-1998 (Rel. 37, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   30S ribosomal protein S13.
GN   RPSM OR BB0500.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: Located at the top of the head of the 30S subunit, it
CC       contacts several helices of the 16S rRNA. In the 70S ribosome it
CC       contacts the 23S rRNA (bridge B1a) and protein L5 of the 50S
CC       subunit (bridge B1b), connecting the 2 subunits; these bridges are
CC       implicated in subunit movement. Contacts the tRNAs in the A and P-
CC       sites (By similarity).
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a loose
CC       heterodimer with protein S19. Forms two bridges to the 50S subunit
CC       in the 70S ribosome (By similarity).
CC   -!- SIMILARITY: Belongs to the S13P family of ribosomal proteins.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001152; AAC66842.1; -.
DR   PIR; C70162; C70162.
DR   TIGR; BB0500; -.
DR   HAMAP; MF_01315; -; 1.
DR   InterPro; IPR001892; Ribosomal_S13.
DR   Pfam; PF00416; Ribosomal_S13; 1.
DR   PROSITE; PS00646; RIBOSOMAL_S13_1; 1.
DR   PROSITE; PS50159; RIBOSOMAL_S13_2; 1.
KW   Ribosomal protein; RNA-binding; rRNA-binding; tRNA-binding;
KW   Complete proteome.
SQ   SEQUENCE   125 AA;  14077 MW;  E51C54E5F791FC95 CRC64;

  Query Match             0.5%;  Score 7;  DB 1;  Length 125;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        503 SGIDLPN 509
              |||||||
Db          5 SGIDLPN 11


Search completed: September 10, 2004, 05:07:24
Job time : 13.2095 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 52.7504 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 1297
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	3.5	50	2	O06019	O06019 clostridium
2	16	1.2	451	2	Q9R631	Q9r631 clostridium
3	16	1.2	1291	2	Q9ZAJ8	Q9zaj8 clostridium
4	16	1.2	1291	2	Q93G71	Q93g71 clostridium
5	16	1.2	1291	2	Q933K0	Q933k0 clostridium
6	16	1.2	1291	2	Q08077	Q08077 clostridium
7	16	1.2	1291	2	Q8GR96	Q8gr96 clostridium
8	15	1.2	361	2	Q45848	Q45848 clostridium
9	15	1.2	361	2	Q45846	Q45846 clostridium
10	14	1.1	1310	2	Q93N27	Q93n27 clostridium
11	13	1.0	1280	2	Q9ZAJ5	Q9zaj5 clostridium
12	12	0.9	1251	2	Q9K395	Q9k395 clostridium
13	12	0.9	1252	2	Q8KZM3	Q8kzm3 clostridium
14	12	0.9	1255	2	Q9FAR6	Q9far6 clostridium
15	12	0.9	1268	2	Q45851	Q45851 clostridium
16	11	0.8	451	2	Q9LA13	Q91la13 clostridium
17	10	0.8	49	2	O87711	O87711 clostridium
18	10	0.8	1275	12	Q9QTG7	Q9qtg7 clostridium
19	10	0.8	1280	2	Q45849	Q45849 clostridium
20	10	0.8	1280	2	Q9LBS7	Q9lbs7 clostridium
21	10	0.8	1280	2	Q841S3	Q841s3 clostridium
22	10	0.8	1285	2	Q45967	Q45967 clostridium
23	10	0.8	1285	2	Q9LBR1	Q9lbr1 clostridium
24	10	0.8	1291	2	Q93HT3	Q93ht3 clostridium
25	9	0.7	441	2	Q9X708	Q9x708 clostridium
26	9	0.7	468	16	Q8XKY8	Q8xky8 clostridium
27	9	0.7	739	16	Q7V1L8	Q7v1l8 prochloroco
28	9	0.7	1148	10	O23603	O23603 arabidopsis
29	8	0.6	92	16	Q81AU7	Q81au7 bacillus ce
30	8	0.6	124	8	Q94TV4	Q94tv4 hapalogaste
31	8	0.6	140	8	Q94TU9	Q94tu9 labidochiru
32	8	0.6	142	8	Q94TU8	Q94tu8 lopholithod
33	8	0.6	146	17	Q9HLW0	Q9hlw0 thermoplasm
34	8	0.6	147	8	Q94TU6	Q94tu6 pagurus ber
35	8	0.6	148	16	Q8ENW4	Q8enw4 oceanobacil
36	8	0.6	155	8	Q94TV7	Q94tv7 cryptolitho
37	8	0.6	155	8	Q36509	Q36509 prodontria
38	8	0.6	156	8	Q36531	Q36531 prodontria
39	8	0.6	159	8	Q8WAC8	Q8wac8 pachycheles
40	8	0.6	159	8	Q8WFX5	Q8wfx5 coenobita c
41	8	0.6	161	12	Q91PR4	Q91pr4 hepatitis b
42	8	0.6	161	12	Q9E991	Q9e991 hepatitis b
43	8	0.6	161	12	Q91PQ8	Q91pq8 hepatitis b
44	8	0.6	161	12	Q91PQ9	Q91pq9 hepatitis b
45	8	0.6	161	12	Q9E990	Q9e990 hepatitis b

ALIGNMENTS

RESULT 1
O06019
ID O06019 PRELIMINARY; PRT; 50 AA.
AC O06019;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BoNT protein (Fragment).
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27322;
RX MEDLINE=97435242; PubMed=9290060;
RA Bhandari M., Campball K.D., Collins M.D., East A.K.;
RT "Molecular characterization of the clusters of genes encoding the
RT botulinum neurotoxin complex in Clostridium Botulinum (clostridium
RT argentinense) type g and non-proteolytic clostridium botuliunum type
RT b.";
RL Curr. Microbiol. 35:207-214(1997).
DR EMBL; X87972; CAA61229.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
FT NON_TER 50 50
SQ SEQUENCE 50 AA; 6036 MW; 04B0316B5A95C199 CRC64;

Query Match 3.5%; Score 45; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 3.4e-38;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWI 45
              |||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWI 45


RESULT 2
Q9R631
ID   Q9R631      PRELIMINARY;      PRT;   451 AA.
AC   Q9R631;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B light chain, BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340509; PubMed=1634516;
RA   Kurazono H., Mochida S., Binz T., Eisel U., Quanz M., Grebenstein O.,
RA   Wernars K., Poulain B., Tauc L., Niemann H.;
RT   "Minimal essential domains specifying toxicity of the light chains of
RT   tetanus toxin and botulinum neurotoxin type A.";
RL   J. Biol. Chem. 267:14721-14729(1992).
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   451 AA;  51943 MW;  6C79FD488653EA71 CRC64;

  Query Match             1.2%;  Score 16;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 3
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match             1.2%;  Score 16;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 4
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             1.2%;  Score 16;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 5
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
```

```
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             1.2%;  Score 16;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 6
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match             1.2%;  Score 16;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 7
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             1.2%;  Score 16;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYGIK 243
              ||||||||||||||||
Db        228 LMHELIHVLHGLYGIK 243


RESULT 8
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;

  Query Match             1.2%;  Score 15;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        708 QWLSTVNTQFYTIKE 722
              |||||||||||||||
Db         70 QWLSTVNTQFYTIKE 84


RESULT 9
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type B;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;

  Query Match             1.2%;  Score 15;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        708 QWLSTVNTQFYTIKE 722
              |||||||||||||||
Db         70 QWLSTVNTQFYTIKE 84


RESULT 10
Q93N27
ID   Q93N27      PRELIMINARY;      PRT;  1310 AA.
AC   Q93N27;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tetanus toxin (Fragment).
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shumin Z., Dianliang L.;
RT   "Cloning and sequence analysis of tetanus toxin gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF389424; AAK72964.2; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER    1310   1310
SQ   SEQUENCE   1310 AA;  150316 MW;  9EADDC914418E450 CRC64;

  Query Match             1.1%;  Score 14;  DB 2;  Length 1310;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        228 LMHELIHVLHGLYG 241
              ||||||||||||||
Db        232 LMHELIHVLHGLYG 245


RESULT 11
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
```

```
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             1.0%;  Score 13;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 0.0012;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         89 LKTMIKLFNRINS 101
              |||||||||||||
Db         88 LKTMIKLFNRINS 100


RESULT 12
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             0.9%;  Score 12;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        223 DPALTLMHELIH 234
              ||||||||||||
Db        205 DPALTLMHELIH 216


RESULT 13
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match              0.9%;  Score 12;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        223 DPALTLMHELIH 234
              ||||||||||||
Db        205 DPALTLMHELIH 216


RESULT 14
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match              0.9%;  Score 12;  DB 2;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        223 DPALTLMHELIH 234
              ||||||||||||
Db        208 DPALTLMHELIH 219


RESULT 15
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match              0.9%;  Score 12;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        230 HELIHVLHGLYG 241
              ||||||||||||
Db        219 HELIHVLHGLYG 230
```

Search completed: September 10, 2004, 05:12:59
Job time : 54.7504 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 85.5545 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 6816
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4409.5	64.7	848	5	ABG69087	Abg69087 Botulinum
2	3992	58.6	1291	2	AAW68392	Aaw68392 Clostridi
3	3684.5	54.1	1169	2	AAW56017	Aaw56017 Recombina
4	2836.5	41.6	858	2	AAW56018	Aaw56018 Recombina
5	2817.5	41.3	852	5	ABG69077	Abg69077 Botulinum
6	2541	37.3	1070	3	AAY93308	Aay93308 A mangane
7	2541	37.3	1095	3	AAY93311	Aay93311 A mangane
8	2538.5	37.2	1315	4	AAB61169	Aab61169 Clostridi
9	2478	36.4	848	4	AAB04082	Aab04082 Botulism
10	2392	35.1	449	3	AAY77139	Aay77139 Synthetic
11	2392	35.1	449	4	AAB04167	Aab04167 Botulism
12	2387	35.0	473	2	AAW68400	Aaw68400 Clostridi
13	2332	34.2	444	5	ABG69086	Abg69086 Botulinum
14	2298	33.7	439	5	ABG69073	Abg69073 Botulinum
15	2285	33.5	1296	2	AAR95010	Aar95010 C. botuli
16	2268	33.3	1295	5	AAU99339	Aau99339 Clostridi
17	2095	30.7	410	4	AAB04102	Aab04102 Botulism
18	1885	27.7	1291	2	AAY05814	Aay05814 Non-toxic
19	1792.5	26.3	548	2	AAW56014	Aaw56014 Recombina
20	1723	25.3	1067	3	AAY93307	Aay93307 A mangane
21	1723	25.3	1092	3	AAY93310	Aay93310 A mangane
22	1699	24.9	837	3	AAY77140	Aay77140 Native bo
23	1690.5	24.8	847	4	AAB04081	Aab04081 Botulism
24	1551.5	22.8	1059	3	AAY93309	Aay93309 A mangane
25	1551.5	22.8	1084	3	AAY93312	Aay93312 A mangane
26	1511	22.2	858	5	ABG69085	Abg69085 Botulinum
27	1502	22.0	953	2	AAW56011	Aaw56011 Recombina
28	1499	22.0	871	2	AAW56008	Aaw56008 Botulinum
29	1497.5	22.0	1013	2	AAW56013	Aaw56013 Recombina
30	1494.5	21.9	907	2	AAW56012	Aaw56012 Recombina
31	1494	21.9	871	2	AAW56019	Aaw56019 Recombina
32	1494	21.9	871	2	AAW56007	Aaw56007 Recombina
33	1494	21.9	873	2	AAW56016	Aaw56016 Recombina
34	1494	21.9	894	2	AAW56015	Aaw56015 Recombina
35	1493.5	21.9	878	2	AAW56010	Aaw56010 Recombina
36	1492	21.9	875	2	AAW56009	Aaw56009 Recombina
37	1473.5	21.6	441	5	ABG69068	Abg69068 Botulinum
38	1473.5	21.6	441	5	ABG69076	Abg69076 Botulinum
39	1473.5	21.6	441	5	ABB80654	Abb80654 Botulinum
40	1473.5	21.6	441	6	ABU61545	Abu61545 Botulinum
41	1472	21.6	804	5	ABG69083	Abg69083 Botulinum
42	1467	21.5	861	5	ABG69075	Abg69075 Botulinum
43	1409	20.7	824	5	ABG69081	Abg69081 Botulinum
44	1366.5	20.0	413	4	AAB04097	Aab04097 Botulism
45	1262	18.5	811	5	ABG69079	Abg69079 Botulinum

ALIGNMENTS

RESULT 1
ABG69087
ID ABG69087 standard; protein; 848 AA.
XX
AC ABG69087;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #21.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US047230.
XX
PR 06-NOV-2000; 2000US-0246774P.
PR 20-JUL-2001; 2001US-00910186.
PR 09-AUG-2001; 2001US-0311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98557.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium.
XX
PS Claim 52; Page 162-164; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein
XX
SQ Sequence 848 AA;

Query Match 64.7%; Score 4409.5; DB 5; Length 848;
Best Local Similarity 99.9%; Pred. No. 1e-254;
Matches 848; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy    1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db    1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy   61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db   61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy  121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db  121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy  181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db  181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy  241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db  241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy  301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
Db  301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy  361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
Db  361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy  421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
Db  421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy  481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
Db  481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy  541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
Db  541 SLFEYLHAQTFPSNIEN-QLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 599

Qy  601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
Db  600 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 659

Qy  661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
Db  660 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 719

Qy  721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
Db  720 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 779

Qy  781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
Db  780 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 839

Qy  841 IPFDLSLYT 849
Db  840 IPFDLSLYT 848
```

RESULT 2
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type B toxin.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30579.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 291-296; 428pp; English.
XX
CC This is the amino acid sequence of the type B toxin of Clostridium
CC botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC B toxin has been expressed as a histidine-tagged protein in Escherichia
CC coli. The invention relates to C. botulinum recombinant toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of the recombinant proteins are E. coli, insect cells and
CC yeast cells. The recombinant proteins are used as immunogens for the
CC production of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
CC (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1291 AA;

Query Match 58.6%; Score 3992; DB 2; Length 1291;
Best Local Similarity 57.6%; Pred. No. 1.6e-229;
Matches 751; Conservative 251; Mismatches 283; Indels 18; Gaps 11;

```
Qy    1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db    1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy   61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db   61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy  121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db  121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy  181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db  181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240
```

```
Qy       241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db       241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy       301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
Db       301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy       360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
Db       361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy       420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
Db       421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy       480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
Db       476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy       540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
Db       535 NTIFQYLYSQTFLLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy       600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
Db       595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy       660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
Db       655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy       720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
Db       715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIHNINNF 774

Qy       780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
Db       775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy       840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
Db       835 IMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN 894

Qy       900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
Db       895 D--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy       960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
Db       953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy      1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
Db      1012 NNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy      1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
Db      1072 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNS 1131

Qy      1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
Db      1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKE 1189

Qy      1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
Db      1190 EEKLFLAPISDSDEFYNTIQIKEYHEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy      1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
Db      1249 VFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

```
RESULT 3
AAW56017
ID   AAW56017 standard; protein; 1169 AA.
XX
AC   AAW56017;
XX
DT   27-JUL-1998  (first entry)
XX
DE   Recombinant botulinum neurotoxin type B LH728/B.
XX
KW   Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW   detection; tetanus; non-toxic; toxin.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO9807864-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-GB002273.
XX
PR   23-AUG-1996;   96GB-00017671.
PR   13-DEC-1996;   96GB-00025996.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA   (SPEY-) SPEYWOOD LAB LTD.
XX
PI   Shone CC,  Quinn CP,  Foster KA;
XX
DR   WPI; 1998-169168/15.
DR   N-PSDB; AAV26289.
XX
PT   Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT   immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS   Example 2; Page 91-94; 137pp; English.
XX
CC   The present sequence represents a recombinant neurotoxin protein from the
CC   present invention. The present invention describes recombinant neurotoxin
CC   proteins which comprise a first and second domain, where the first domain
CC   is adapted to cleave one or more vesicle or plasma-membrane associated
CC   proteins essential to exocytosis, and where the second domain is adapted:
CC   (a) to translocate the protein into a cell; (b) to increase the
CC   solubility of the protein compared to the solubility of the first domain
CC   on its own, or (c) both to translocate the protein into a cell and to
CC   increase the solubility of the protein compared to the solubility of the
CC   first domain on its own, the protein being free of clostridial neurotoxin
CC   (CN) and free of CN precursor that can be converted into toxin by
CC   proteolytic action. The recombinant proteins can be used as therapeutic
CC   agents for targeting cells expressing a relevant substrate. The products
CC   can also be used as immunogens and as non-toxic standards for the
CC   assessment and development of in vitro assays for the detection of
CC   functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC   environmental samples
XX
SQ   Sequence 1169 AA;
```

```
  Query Match             54.1%;  Score 3684.5;  DB 2;  Length 1169;
  Best Local Similarity   58.7%;  Pred. No. 3.2e-211;
  Matches  692;  Conservative  226;  Mismatches  248;  Indels   13;  Gaps    8;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180
```

```
Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
Db        476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
Db        595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
Db        835 IMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
Db        895 D--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
Db        953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy       1020 GNANIYIINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
Db       1012 NNAKIYIINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
Db       1072 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYL 1176
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYL 1169
```

RESULT 4
AAW56018
ID AAW56018 standard; protein; 858 AA.

XX
AC AAW56018;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH417/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
KW detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB002273.
XX
PR 23-AUG-1996; 96GB-00017671.
PR 13-DEC-1996; 96GB-00025996.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Shone CC, Quinn CP, Foster KA;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26290.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
PT immunogens or as non-toxic standards for the detection of neurotoxins.
XX
PS Example 2; Page 98-100; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from the
CC present invention. The present invention describes recombinant neurotoxin
CC proteins which comprise a first and second domain, where the first domain
CC is adapted to cleave one or more vesicle or plasma-membrane associated
CC proteins essential to exocytosis, and where the second domain is adapted:
CC (a) to translocate the protein into a cell; (b) to increase the
CC solubility of the protein compared to the solubility of the first domain
CC on its own, or (c) both to translocate the protein into a cell and to
CC increase the solubility of the protein compared to the solubility of the
CC first domain on its own, the protein being free of clostridial neurotoxin
CC (CN) and free of CN precursor that can be converted into toxin by
CC proteolytic action. The recombinant proteins can be used as therapeutic
CC agents for targeting cells expressing a relevant substrate. The products
CC can also be used as immunogens and as non-toxic standards for the
CC assessment and development of in vitro assays for the detection of
CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
CC environmental samples
XX
SQ Sequence 858 AA;

Query Match 41.6%; Score 2836.5; DB 2; Length 858;
Best Local Similarity 61.6%; Pred. No. 9.2e-161;
Matches 532; Conservative 157; Mismatches 168; Indels 7; Gaps 4;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
```

```
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||: :||| || |:|||| :| :|||||||||| |||:|:|||| :||:| ||||: |
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:|||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| |||: |||:| : |||| ||| |||||: |::: |: |||||:||:|
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  | |    || |:|||||||:|:||| ||:| | | |
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              |||:| |||:| |::|||| || | |:||:|||| |:| ::|:||| || |:|||| |
Db        476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | :::||  |: : :|:||::|:|:|:|||| |||||||||| ||:||||| ||:||:||
Db        595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | |||||: ||  || || ||| ||::||:|||||||:||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              |||.|| ||| |:||:|:||: :|| |||::| |||||||||: |||: || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              || ||:|||| :||||||:|| ||||: ||::|| ||| |:|||:      |||||::||
Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYIS 863
              :|||||:|| |||||::|| | |
Db        835 IMPFDLSIYTNDTILIEMFNKYNS 858


RESULT 5
ABG69077
ID   ABG69077 standard; protein; 852 AA.
XX
AC   ABG69077;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #11.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98547.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 52; Page 138-139; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 852 AA;

  Query Match             41.3%;  Score 2817.5;  DB 5;  Length 852;
  Best Local Similarity   61.5%;  Pred. No. 1.2e-159;
  Matches  528;  Conservative  157;  Mismatches  166;  Indels    7;  Gaps    4;

Qy          2 PVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNA 61
              || | |||||||||:|::|||||     | | |||||:| |||||:|||:|:|::|: ||
Db          1 PVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNK 60

Qy         62 STGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGN 121
              |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::||:||:: |||||:
Db         61 SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLGD 120

Qy        122 ASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHS 181
                | ::|  |:|:|::|| |  ||  :: ||:  |||||||||||::| |  : :  |
Db        121 RRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHF 180

Qy        182 PISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYG 241
                ||||  | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||||
Db        181 ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLYG 240

Qy        242 IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIAN 301
              ||: :||| || |:|||| :| :|||||||||| |||:|:|||| :||:| ||||: | :
Db        241 IKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIVD 300

Qy        302 RLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:|||||||||:
Db        301 RLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETNI 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              |  | |||| |||: |||:| : |||| ||| ||||: |::: |: |||||:||:||
Db        361 AENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              |||| ||| :|:| ||| |  |  |    || |:|||||||:|:||| ||:| | | ||
Db        421 EEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEYN 475

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||:| |||:| |::|||| || | |:||:|||| |:| ::|:||| || |:|||| | :
Db        476 TQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDEN 534

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
```

```
                ::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  |||
Db        535 TIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVK 594

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              :::||  |: : :|:||::|:|:|:|||| |||||||||| ||:||||| ||:||:|||
Db        595 QIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEFI 654

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              |||::|:|| | |||||: ||  || || ||| ||::||:|||||||:|||||||||||||
Db        655 PELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTI 714

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              || || ||| |:||:|:||: :|| |||::| |||||||||: |||: || ||:||::||
Db        715 KEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNFI 774

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              | ||:|||| :||||||:|| |||: ||::|| ||| |:|||:       |||||::||
Db        775 NGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTI 834

Qy        841 IPFDLSLYTKDTILIQVF 858
              :|||||:|| |||||::|
Db        835 MPFDLSIYTNDTILIEMF 852
```

RESULT 6
AAY93308
ID AAY93308 standard; protein; 1070 AA.
XX
AC AAY93308;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype B.
XX
OS Synthetic.
OS Geobacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB003699.
XX
PR 05-NOV-1998; 98GB-00024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a cleavable
PT linker to a neuronal cell targeting component useful for delivering
PT superoxide dismutase to neuronal cells to treat ischemia.
XX
PS Disclosure; Page 45-47; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype B. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a domain
CC that translocates the SOD of the composition into the neuronal cell.
CC After translocation, the linker is cleaved to release the SOD. The
CC composition is useful for treating neuronal diseases caused or augmented
CC by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC disease, Huntington's disease and motor neurone diseases
XX
SQ Sequence 1070 AA;

Query Match 37.3%; Score 2541; DB 3; Length 1070;
Best Local Similarity 56.4%; Pred. No. 5.5e-143;
Matches 481; Conservative 173; Mismatches 187; Indels 12; Gaps 8;

```
Qy        450 CIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPN 509
              || |:|||||||:|:||| ||:| | | ||||:| |||:| |::|||| || | |:||:
Db        225 CIDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELILDTDLISKIELPS 284

Qy        510 ENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALR 569
              ||||  |:| ::|:||| || |:|||| | :::|:||::|||| :| :: ||:| :|||
Db        285 ENTESLTDF-NVDVPVYEKQPAIKKIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALL 343

Qy        570 NNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYI 629
              :||||:||| : :: || || | ||  ||| :::||  |: : :|:||::|:|:|:|||
Db        344 FSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYI 403

Qy        630 GPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISN 689
              | |||||||||| ||:||||| ||:||:||||||::|:|| | ||||: ||  || || | 
Db        404 GLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIKTIDN 463

Qy        690 ALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEE 749
              || ||::||:|||||||:|||||||||||||| || ||| |:||:|:||: :|| |||:
Db        464 ALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEK 523

Qy        750 DKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKR 809
              :| |||||||||: |||: || ||:||::||| ||:|||| :|||||:|| |||: ||:
Db        524 EKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKK 583

Qy        810 DLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNA 869
              :|| ||| |:|||:       |||||::||  :|||||:|| |||||::|| | | | :| 
Db        584 NLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNI 643

Qy        870 ILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSM 929
              ||:| |:   ||| ||||| : |   |  ||    |||| :| || |   |:: ::::|:
Db        644 ILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSANSKIRVTQNQNIIFNSV 701

Qy        930 FDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKS 989
              | :||::||:| ||| |: || |: |||||:|:||:||||:||:|||||||||:| |:
Db        702 FLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKT 761

Qy        990 KSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDID 1049
              ||:||||:|:::||:|||:|| :||||: | || ||||| |: :  | ::  : :: :| 
Db        762 KSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEII 820

Qy       1050 FKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLF 1109
              |||    | |:|:|:| |:||  ||: : :   | ||| :  |||||||||| |: :||:|
Db        821 FKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMF 880

Qy       1110 NQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLRFIIKKASNSRNINNDNI 1167
              | | :| |||    : :||   |: :|  :  |||::||:| :|||:: |||::| ||:|
Db        881 NAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSI-NDDI 939

Qy       1168 VREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTY 1227
              ||: |||||:  |:: | :|||     |: : :||||||:|   ||: :|||:| |: ||
Db        940 VRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTY 998

Qy       1228 NCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGC 1284
              :||:| :||    |   || || :| :  | |:: | :|||||:|||: :      |:|||
Db        999 SCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCISKWYLKEVKRKPYNLKLGC 1057

Qy       1285 NWQFIPVDEGWTE 1297
              |||||| ||||||
Db       1058 NWQFIPKDEGWTE 1070
```

RESULT 7
AAY93311

```
ID   AAY93311 standard; protein; 1095 AA.
XX
AC   AAY93311;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype B.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 54-57; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide dismutase
CC   (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC   cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC   serotype B. The specification describes a composition for delivery of SOD
CC   to neuronal cells. The composition comprises SOD linked, by a cleavable
CC   linker, to a neuronal cell targeting component (NCTC). This component has
CC   a domain that binds to a neuronal cell and a domain that translocates the
CC   SOD of the composition into the neuronal cell. After translocation, the
CC   linker is cleaved to release the SOD. The composition is useful for
CC   treating neuronal diseases caused or augmented by oxidative stress, such
CC   as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases
XX
SQ   Sequence 1095 AA;

  Query Match             37.3%;  Score 2541;  DB 3;  Length 1095;
  Best Local Similarity   56.4%;  Pred. No. 5.7e-143;
  Matches  481;  Conservative  173;  Mismatches  187;  Indels   12;  Gaps    8;

Qy        450 CIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPN 509
              || |:|||||||:|:||| ||:| | | ||||:| |||:| |::|||| || | |:||:
Db        250 CIDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELILDTDLISKIELPS 309

Qy        510 ENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALR 569
              ||||  |:| ::|:||| || |:|||| | :::|:||::|||| :| :: ||:| :|||
Db        310 ENTESLTDF-NVDVPVYEKQPAIKKIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALL 368

Qy        570 NNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYI 629
               :||||:||| : :: || || | ||  ||| :::||  |: : :|:||::|:|:|:||| 
Db        369 FSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYI 428

Qy        630 GPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISN 689
              | |||||||||| ||:||||| ||:||:||||||::|:|| | ||||: ||  || || |

Db        429 GLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIKTIDN 488

Qy        690 ALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEE 749
              || ||::||:|||||||:|||||||||||||| || ||| |:||:|:||: :|| |||:
Db        489 ALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEK 548

Qy        750 DKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKR 809
              :| |||||||||: |||: || ||:||::||| ||:|||| :|||||:|| |||: ||:
Db        549 EKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKK 608

Qy        810 DLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNA 869
              :|| ||| |:|||:       |||||::||  :|||||:|| ||||::|| | | | :|
Db        609 NLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNI 668

Qy        870 ILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSM 929
              ||:| |:   ||| ||||| : |   |  ||    |||| :| || |   |:: ::::|:
Db        669 ILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSANSKIRVTQNQNIIFNSV 726

Qy        930 FDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKS 989
              | :||::||:| ||| |: || |: |||||:|:|:||:||||:||:|||||||||:| |:
Db        727 FLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKT 786

Qy        990 KSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDID 1049
              ||:||||:|:::||:|||:|| :|||| | || ||||| |: :  | ::  : :: :|
Db        787 KSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEII 845

Qy       1050 FKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLF 1109
              |||   | |:|:|:| |:|| ||: : :   | ||| :  |||||||| |: :||:|
Db        846 FKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMF 905

Qy       1110 NQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLRFIIKKASNSRNINNDNI 1167
              | | :| |||    : :||   |: :|  :  |||::||:| :|||:: |||::| ||:|
Db        906 NAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSI-NDDI 964

Qy       1168 VREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTY 1227
              ||: |||||:  |:: | :|||     |: : :|||||:|   ||: :|||:| |: ||
Db        965 VRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISDSDEFYNTIQIKEYDEQPTY 1023

Qy       1228 NCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGC 1284
              :||:| :||   |   || || :| :  | |:: | :|||||:|||: :      |:|||
Db       1024 SCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCISKWYLKEVKRKPYNLKLGC 1082

Qy       1285 NWQFIPVDEGWTE 1297
              |||||| ||||||
Db       1083 NWQFIPKDEGWTE 1095


RESULT 8
AAB61169
ID   AAB61169 standard; protein; 1315 AA.
XX
AC   AAB61169;
XX
DT   02-APR-2001  (first entry)
XX
DE   Clostridium tetani TeNT.
XX
KW   Clostridium tetani; TeNT; tetanus toxin; antibacterial; vaccine;
KW   TeNT fragment C; infection.
XX
OS   Clostridium tetani.
XX
PN   WO200100839-A1.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-GB002428.
XX
PR   25-JUN-1999;   99GB-00014861.
XX
PA   (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI   Fairweather NF,  Sinha K;
```

XX
DR WPI; 2001-123014/13.
XX
PT New polypeptide, useful for treating infections of Clostridium tetani,
PT comprises tetanus toxin fragment with a mutation in a loop region,.
XX
PS Disclosure; Page 39; 43pp; English.
XX
CC The present sequence is given in a specification relating to a novel
CC polypeptide comprising tetanus toxin (TeNT) fragment C or its immunogenic
CC fragment, containing a mutation in a loop region. The mutation results in
CC a reduction in the binding of TeNT fragment C or its immunogenic fragment
CC to gangliosides and primary motoneurones, and/or a reduction in the
CC ability of TeNT fragment C or its immunogenic fragment to undergo
CC retrograde transport. The polypeptide is useful for treating, preventing
CC and reducing the susceptibility to Clostridium tetani infection in a
CC human or animal, and also for producing antibodies which recognise groups
CC within TeNT polypeptides. Antibody produced against the polypeptide is
CC also useful for treating Clostridium tetani infection
XX
SQ Sequence 1315 AA;

Query Match 37.2%; Score 2538.5; DB 4; Length 1315;
Best Local Similarity 40.1%; Pred. No. 1e-142;
Matches 543; Conservative 252; Mismatches 460; Indels 99; Gaps 27;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||: | || |:||:||| ||||||    |   |||||:| |||||||||: :| :|: ||
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                : :  :   ||||| ||:||::||:||:||:|||||| :  :|: ||| |::||||||
Db         61 PPSSLI-EGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLG 119

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM--- 177
              |: :  |||  |  :|| |     |          ::||||||||||||: |    :::
Db        120 NSYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVD 179

Qy        178 -NGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVL 236
               : |  :|||: | : |||  :  |:|| ||  :    :  || |||| |||||||||
Db        180 NKNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVL 239

Qy        237 HGLYGIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNF 296
              |||||:::|:  | |: :| :|||: |: ||||:|||| | ::||     ::| | | ::
Db        240 HGLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDY 299

Qy        297 QDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFT 356
               : |||:|: |:|   |||  |||||: || | :| ||:| |::|||  || ::|:|||
Db        300 KAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFT 359

Qy        357 ETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVN 416
              |  |  :: |||| ||||    |:|   |||:|||   ||||| ||:||:|: |||  ||
Db        360 EIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVN 419

Qy        417 KEAYEEISLEHLVIYRIAMCKPV---------MYKNTGK-----SEQCIIVNNEDLFFIA 462
                |:  :    ||   | :|| :          :|  |         | || : |||| |||
Db        420 TNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFIA 479

Qy        463 NKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDID 522
              |:|||::  : | ::|||:|  :  |:|:|::|:| :| | | |||: | | |    |
Db        480 EKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTK--GIP 537

Qy        523 IPVYIKQSALKKI---FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFS 579
                  | :|   |    :| :::::||:||  |: :  : :|||::||| |: |:|::|
Db        538 YAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLGRITMTNSVDDALINSTKIYSYF- 596

Qy        580 TNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNET 639
               ::: | |      ||: ||: :|||||:||:|:|||||:|||| |:|||||||:  :
Db        597 PSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQG 656

Qy        640 AKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWT 699
              : ||  | | |  :|:|:|||: :|::   ::       |  || || | |:|| :||
Db        657 YEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWI 716

Qy        700 DMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFN 759
              ::| |: ::|| ||||||     :|| :|  |  ||:|||: :|  ||  ||  |  : |
Db        717 EVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEIN 776

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819
              ::  || :  | |: ||: |: : | |:|:|:||  | |:| :||   |  |::|| ' |
Db        777 NLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANS 836

Qy        820 LYL-LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN-------ISSNAIL 871
              :: : |:  |:||:|:     |||    |:|:       : :: |       :  : ||
Db        837 KFIGITELKKLESKINKVFSTPIPFS---YSKN------LDCWVDNEEDIDVILKKSTIL 887

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMF 930
              :|      : | ||: :::    |  :    |      | |:|:|  :  |::  : |: ||
Db        888 NLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMF 947

Qy        931 DNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRIIWTLIDV 985
              :||:::||:| || : : :: |  |||:||| :|       ||| ||:||| :|||| |
Db        948 NNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTLKDS 1007

Qy        986 NAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINS 1044
               : | |  :  | :  | :|: |||  |||||||| :||:|||| |  | :|  |  |
Db       1008 AGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIRE 1066

Qy       1045 SNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDT 1104
              |:|  ||  | :  ::| |  | || : ||  |:  ||     |  |:|||||||||||
Db       1067 DNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDT 1126

Qy       1105 QYYLF-------NQGMQNI--YIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFII 1153
              :|||         : |::||  |: | : |         :: |  :|  |: || ||:|||
Db       1127 EYYLIPVASSSKDVGLKNITDYM-YLTNA--------PSYTNGKLNIYYRRLYNGLKFII 1177

Qy       1154 KKASNSRNINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPIND 1208
              |: : :  |  |: |: ||:| | :   ::|     |          |   |    : ||
Db       1178 KRYTPNNEI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP--- 1232

Qy       1209 DPTFYDVLQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNY 1262
                |  ::  |  :  ||: |: | :    : || |     ||          |   :
Db       1233 GIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDI 1283

Qy       1263 FCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 1296
                | ||   : : |    |||:| |:| |||||
Db       1284 LIASNWYFNHLKDKI----LGCDWYFVPTDEGWT 1313
```

RESULT 9
AAB04082
ID AAB04082 standard; protein; 848 AA.
XX
AC AAB04082;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.

PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Page 13; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 848 AA;

Query Match 36.4%; Score 2478; DB 4; Length 848;
Best Local Similarity 56.0%; Pred. No. 2.3e-139;
Matches 478; Conservative 171; Mismatches 189; Indels 16; Gaps 12;

```
Qy        450 CIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPN 509
              || |:||||||||:|:|||| ||:|  | | |||||:| |||:|  |::|||| || | |:||:
Db          5 CIDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPINELILDTDLISKIELPS 64

Qy        510 ENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALR 569
              ||||  |:| ::|:||| || |:|||| | :::|:||::|||| :| :: ||:| :|||
Db         65 ENTESLTDF-NVDVPVYEKQPAIKKIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALL 123

Qy        570 NNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYI 629
              :||||:||| : :: || || | ||  ||| :::||  |: : :|:||::|:|:|:|||
Db        124 FSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYI 183

Qy        630 GPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISN 689
              | |||||||||| ||:||||| ||:||:||||||::|:|| | ||||: ||  || || |
Db        184 GLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLLESYIDNKNKIIKTIDN 243

Qy        690 ALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEE 749
              || ||::||:|||||||:|||||||||||||| || ||| |:||:|:||: :|| |||:
Db        244 ALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEK 303

Qy        750 DKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKR 809
              :| |||||||||: |||: || ||:||::||| ||:|||| :|||||:|| |||: ||:
Db        304 EKSNINIDFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKK 363

Qy        810 DLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNA 869
              :|| ||| |:|||:      |||||::||  :|||||:|| ||||||::|| | | | :|
Db        364 NLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSEILNNI 423

Qy        870 ILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSM 929
              ||:| |:   ||| ||||| : |   | ||    |||| :| || |   |:: :::.:|:
Db        424 ILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSANSKIRTTQNQNIIFNSV 481

Qy        930 FDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKS 989
              | :||::||:| ||| |: || |: |||||:|:||:||||:||:| |||||||:| |:
Db        482 FLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRG-RIIWTLIDINGKT 540

Qy        990 KSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDID 1049
              ||:||||:|:::||:|||:|| :||||: | || ||||| |: :  | ::  : :: :|
Db        541 KSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKDIREVIANGEII 599

Qy       1050 FKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLF 1109
              |||    | |:|:|:| |:||  ||: : :   | ||| :  |||||||| |: :||:|
Db        600 FKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMF 659

Qy       1110 NQGMQNIY-IKYFSKASMGETAPRTNFNNAA--INYQNLYLGLRFIIKKASNSRNINNDN 1166
              | | :| | ||    : :||   |: :|  :  |||::||:| :|||:: |||::| ||:
Db        660 NAGNKNSYTIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSI-NDD 718

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTT 1226
              |||: |||||:  |:: | :|||     |: : :||||||:|   ||: :|||:| |: |
Db        719 IVRKEDYIYLDFFNLNQE-WRVYTYKYFKK-EEELFLAPISDSDEFYNTIQIKEYDEQPT 776

Qy       1227 YNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLG 1283
              |:||:| :||   |   || || :| :  | |::   : |||| ||| :      |:||
Db        777 YSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEVKDVFCISDWYL-EVKRKPYNLKLG 834

Qy       1284 CNWQFIPVDEGWTE 1297
              ||||||| ||||||
Db        835 CNWQFIPKDEGWTE 848
```

RESULT 10
AAY77139
ID AAY77139 standard; protein; 449 AA.
XX
AC AAY77139;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype G (BoNTG) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype G; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87217.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 28; Page 47-48; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce

CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 449 AA;

Query Match 35.1%; Score 2392; DB 3; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.3e-134;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        850 KDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLN 909
              ||||||
```

```
Qy       1270 LRRISENINKLRLGCNWQFIPVDEGWTE 1297
             ||||||||||||||||||||||||||||||
Db        422 LRRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 12
AAW68400
ID   AAW68400 standard; protein; 473 AA.
XX
AC   AAW68400;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW   botulism; BotG.
XX
OS   Clostridium botulinum; serotype G strain 113/30.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30596.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC   sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC   express BotG soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match             35.0%;  Score 2387;  DB 2;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-134;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Qy        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146

Qy        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206

Qy       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266

Qy       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326

Qy       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386

Qy       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446

Qy       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||
Db        447 RRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 13
ABG69086
ID   ABG69086 standard; protein; 444 AA.
XX
AC   ABG69086;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #20.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98556.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 160-161; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
```

```
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 444 AA;

  Query Match             34.2%;  Score 2332;  DB 5;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 5e-131;
  Matches  441;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMY 441
              |||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMY 441


RESULT 14
ABG69073
ID   ABG69073 standard; protein; 439 AA.
XX
AC   ABG69073;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Botulinum neurotoxin light chain polypeptide #7.
XX
KW   Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW   spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW   bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW   cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW   lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW   immunostimulant; antibacterial.
XX
OS   Clostridium botulinum.
XX
PN   WO200236758-A2.
XX
PD   10-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US047230.
XX
PR   06-NOV-2000; 2000US-0246774P.
PR   20-JUL-2001; 2001US-00910186.
PR   09-AUG-2001; 2001US-0311966P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Jensen M;
XX
DR   WPI; 2002-575192/61.
DR   N-PSDB; ABK98543.
XX
PT   Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT   serotype A, useful for producing the neurotoxin for vaccination against
PT   botulism, comprises sequence expressible in host other than Clostridium.
XX
PS   Claim 33; Page 129-130; 166pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a botulinum
CC   neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC   that is expressible in a host organism other than Clostridium, or has a
CC   total A+T content that is less than about 70% The BoNT LC protein is
CC   useful in vaccination against botulism, for eliciting protective immunity
CC   in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC   facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC   myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC   conditions characterised by hyperactivity of the lower motor neuron, and
CC   to control autonomic nerve function or tiptoe-walking due to stiff
CC   muscles common in children with cerebral palsy. The sequences are also
CC   useful for screening for botulinum neurotoxin inhibitors. This sequence
CC   represents a botulinum neurotoxin light chain serotype A protein
XX
SQ   Sequence 439 AA;

  Query Match             33.7%;  Score 2298;  DB 5;  Length 439;
  Best Local Similarity   99.3%;  Pred. No. 5.3e-129;
  Matches  438;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDD-IMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 59

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 119

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 179

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 239

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        240 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSV-SPSTDMNIYNKALQNFQDIA 298

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 358

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        359 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKNEFNGQNKAVNKEAY 418

Qy        421 EEISLEHLVIYRIAMCKPVMY 441
              |||||||||||||||||||||
Db        419 EEISLEHLVIYRIAMCKPVMY 439


RESULT 15
AAR95010
ID   AAR95010 standard; protein; 1296 AA.
```

XX
AC AAR95010;
XX
DT 09-JUL-1996 (first entry)
XX
DE C. botulinum type A neurotoxin.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.
XX
OS Clostridium botulinum.
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29244.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 4; Page 344-350; 434pp; English.
XX
CC Clostridium botulinum type A neurotoxin (AAR95010) is processed to form a
CC dimer composed of a light and a heavy chain. It is the product of the
CC type A neurotoxin gene (AAT29244). The 50 kDa C-terminal portion of the
CC heavy chain, or C fragment (see also AAR95008), was produced using a
CC synthetic gene (AAT29245) having codon usage altered to improve
CC expression in Escherichia coli. Fusion proteins of the type A toxin or C
CC fragment, with e.g. maltose binding protein or polyhistidine affinity tag
CC (see also AAR95008), are used to generate neutralising antitoxins and in
CC vaccine compsns
XX
SQ Sequence 1296 AA;

Query Match 33.5%; Score 2285; DB 2; Length 1296;
Best Local Similarity 39.7%; Pred. No. 1.4e-127;
Matches 530; Conservative 216; Mismatches 499; Indels 90; Gaps 29;

```
Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
                :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       |  ||      |||    : : |  ||:| ||    :|          ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |      |     |      :     : |||:|| ||||| | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::       | ||| ||||||   |    :    : |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
              |:|||: ||   |  |:   :   | ::| ||   || :||:|||| ||||||| |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
              :|   : ::         |::   : ::           |   :: || ||| ||||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
              ::|:|: || | | | :|     | | |:| ||    |:   |    ||         ||
Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
                | |  |           ||  :|  ::| || || |     : ||||:|:|| |
Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
              ::||||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::||||||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
              |||:||   |::|  |    || ||:|||||: :|::| | | ||: ||   : || |||
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
              ||::|| ::|  ||: ||: ||||    |:::|  || ||::| : ||  |||:|:||:|
Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
              ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:||| :||  |
Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871
              |:||  |   |: :|: || |||  |  ||| || |  :  |:  |  || || : :||
Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSIL 879

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
              :| |    ||| | | : :|:|| | |: |   | :| | |:| |      :||:||::
Db        880 NLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYE 939

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKS 991
              ||| :||:| ||| |:      | ||||||:|::|:|||||:    ||||| |      :
Db        940 NFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR 996

Qy        992 IFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK 1051
              : |:||   |||||||:|  :||||:|| |: ||||| |   : | ||  |::||:| ||
Db        997 VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFK 1056

Qy       1052 LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQ 1111
              |  | || :::||| ||:| :|||  |:  ||  ||::  ||||||: |:|| ||: |
Db       1057 LDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNL 1116

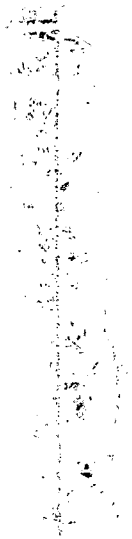
Qy       1112 GMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
                | |:   :  :  |    || :    | :|  :|| | :|||| ::    | ||
Db       1117 YDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDN 1172

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKT 1225
              |||  | :|:|:  : :: ||:    :   ::  |    | |       |::| |  :
Db       1173 IVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGI 1231

Qy       1226 TYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRL 1282
              |  |:: | : :     |  |  :|           :|      | || |:|    :  |
Db       1232 TNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTL 1278

Qy       1283 GCNWQFIPVDEGWTE 1297
                ||:|:|||||:|| |
Db       1279 GCSWEFIPVDDGWGE 1293
```

Search completed: September 10, 2004, 03:33:46
Job time : 88.5545 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 27.5291 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 6816
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3684.5	54.1	1169	4	US-09-255-829-20	Sequence 20, Appl
2	2836.5	41.6	858	4	US-09-255-829-22	Sequence 22, Appl
3	2836.5	41.6	858	4	US-09-255-829-29	Sequence 29, Appl
4	2540.5	37.3	1315	4	US-08-913-880C-1	Sequence 1, Appli
5	2285	33.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
6	2285	33.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
7	2285	33.5	1296	3	US-08-915-136-28	Sequence 28, Appl
8	2285	33.5	1296	4	US-09-084-517-28	Sequence 28, Appl
9	1792.5	26.3	548	4	US-09-255-829-24	Sequence 24, Appl
10	1502	22.0	953	4	US-09-255-829-14	Sequence 14, Appl
11	1499	22.0	871	4	US-09-255-829-8	Sequence 8, Appli
12	1497.5	22.0	1013	4	US-09-255-829-18	Sequence 18, Appl
13	1494.5	21.9	907	4	US-09-255-829-16	Sequence 16, Appl
14	1494	21.9	871	4	US-09-255-829-2	Sequence 2, Appli
15	1494	21.9	871	4	US-09-255-829-26	Sequence 26, Appl
16	1494	21.9	873	4	US-09-255-829-6	Sequence 6, Appli
17	1494	21.9	894	4	US-09-255-829-4	Sequence 4, Appli
18	1493.5	21.9	878	4	US-09-255-829-12	Sequence 12, Appl
19	1492	21.9	875	4	US-09-255-829-10	Sequence 10, Appl
20	1432.5	21.0	865	4	US-08-913-880C-13	Sequence 13, Appl
21	1432.5	21.0	866	4	US-08-913-880C-12	Sequence 12, Appl
22	1432.5	21.0	874	4	US-08-913-880C-11	Sequence 11, Appl
23	1432.5	21.0	875	4	US-08-913-880C-10	Sequence 10, Appl
24	1431	21.0	853	4	US-08-913-880C-17	Sequence 17, Appl
25	1431	21.0	858	4	US-08-913-880C-16	Sequence 16, Appl
26	1431	21.0	860	4	US-08-913-880C-15	Sequence 15, Appl
27	1431	21.0	862	4	US-08-913-880C-14	Sequence 14, Appl
28	815	12.0	438	1	US-08-480-604A-23	Sequence 23, Appl
29	815	12.0	438	2	US-08-405-496A-23	Sequence 23, Appl
30	815	12.0	438	3	US-08-915-136-23	Sequence 23, Appl
31	815	12.0	438	4	US-09-084-517-23	Sequence 23, Appl
32	815	12.0	462	1	US-08-480-604A-26	Sequence 26, Appl
33	815	12.0	462	2	US-08-405-496A-26	Sequence 26, Appl
34	815	12.0	462	3	US-08-915-136-26	Sequence 26, Appl
35	815	12.0	462	4	US-09-084-517-26	Sequence 26, Appl
36	647.5	9.5	618	1	US-08-668-381A-5	Sequence 5, Appli
37	643.5	9.4	452	1	US-07-618-312A-2	Sequence 2, Appli
38	643.5	9.4	452	1	US-08-280-228-2	Sequence 2, Appli
39	642.5	9.4	452	1	US-07-618-312A-4	Sequence 4, Appli
40	642.5	9.4	452	1	US-08-280-228-4	Sequence 4, Appli
41	622.5	9.1	452	1	US-08-110-786A-8	Sequence 8, Appli
42	377	5.5	207	3	US-08-446-114A-21	Sequence 21, Appl
43	279	4.1	2184	4	US-09-417-485D-6	Sequence 6, Appli
44	273.5	4.0	2710	1	US-08-480-604A-6	Sequence 6, Appli
45	273.5	4.0	2710	2	US-08-405-496A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 54.1%; Score 3684.5; DB 4; Length 1169;
Best Local Similarity 58.7%; Pred. No. 1.5e-237;
Matches 692; Conservative 226; Mismatches 248; Indels 13; Gaps 8;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
Db        476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
Db        595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
Db        835 IMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
Db        895 D--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
Db        953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy       1020 GNANIYIYNGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
Db       1012 NNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQPIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
Db       1072 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYL 1176
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYL 1169
```

RESULT 2
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-22

Query Match 41.6%; Score 2836.5; DB 4; Length 858;
Best Local Similarity 61.6%; Pred. No. 4e-181;
Matches 532; Conservative 157; Mismatches 168; Indels 7; Gaps 4;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240
```

```
Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||: :||| || |:|||| :| :|||||||||| |||:|:|||| :||:| |||| |
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:|||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| |||| ||:| : |||| ||| ||||| |::: |: |||||:||:|
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  |  |    || |:|||||||:|:||| ||:| | | |
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              |||:| |||:| |::|||| || | |:||:|||  |:| ::|:|||\|| |:|||| |
Db        476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | :::||  |: : :|:||::|:|:|:|||| |||||||||| ||:||||| ||:||:||
Db        595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | |||: ||  || || ||| ||::||:|||||||:|||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::| |||||||||: |||: || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              || ||:|||| :||||||:|| |||: ||::|| ||| |:|||:      |||||::||
Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYIS 863
               :|||||:|| |||||::|| | |
Db        835 IMPFDLSIYTNDTILIEMFNKYNS 858
```

```
RESULT 3
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  858 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-29

  Query Match             41.6%;  Score 2836.5;  DB 4;  Length 858;
  Best Local Similarity   61.6%;  Pred. No. 4e-181;
  Matches  532;  Conservative  157;  Mismatches  168;  Indels    7;  Gaps    4;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | ||||||||:|::|||||    | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
               |:|:|::|| ||||| || |: :|: ||:|||||||||| ||| |::||:||:: |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::|  |:|:|::|| |  ||  :: ||:  |||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                  ||||  | ::||| ::|||||||||  |||:|| ||:|||| ||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||: :||| || |:|||| :| :|||||||||| |||:|:|||| :||:| |||| |
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:|||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| |||| ||:| : |||| ||| ||||| |::: |: |||||:||:|
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  |  |    || |:|||||||:|:||| ||:| | | |
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              |||:| |||:| |::|||| || | |:||:|||  |:| ::|:|||\|| |:|||| |
Db        476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | :::||  |: : :|:||::|:|:|:|||| |||||||||| ||:||||| ||:||:||
Db        595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | |||: ||  || || ||| ||::||:|||||||:|||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::| |||||||||: |||: || ||:||::|
```

```
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839

Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYIS 863

Db        835 IMPFDLSIYTNDTILIEMFNKYNS 858


RESULT 4
US-08-913-880C-1
; Sequence 1, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
;  APPLICANT: MATSUDA, Morihiro
;  TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
;  TITLE OF INVENTION:  VACCINE
;  FILE REFERENCE: 216-380P
;  CURRENT APPLICATION NUMBER: US/08/913,880C
;  CURRENT FILING DATE:  1997-11-24
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
;   LENGTH: 1315
;   TYPE: PRT
;   ORGANISM: Clostridium tetani
US-08-913-880C-1

  Query Match             37.3%;  Score 2540.5;  DB 4;  Length 1315;
  Best Local Similarity   40.1%;  Pred. No. 4.3e-161;
  Matches  540;  Conservative  253;  Mismatches  469;  Indels   85;  Gaps   25;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db         61 PPSSLI-EGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLG 119

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM--- 177

Db        120 NSYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVD 179

Qy        178 -NGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVL 236

Db        180 NKNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVL 239

Qy        237 HGLYGIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNF 296

Db        240 HGLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDY 299

Qy        297 QDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFT 356

Db        300 KAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFT 359

Qy        357 ETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVN 416

Db        360 EIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVN 419

Qy        417 KEAYEEISLEHLVIYRIAMCKPV---------MYKNTGK-----SEQCIIVNNEDLFFIA 462

Db        420 TNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFIA 479

Qy        463 NKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDID 522

Db        480 EKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTK--GIP 537

Qy        523 IPVYIKQSALKKI---FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFS 579

Db        538 YAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYF- 596

Qy        580 TNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNET 639

Db        597 PSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQG 656

Qy        640 AKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWT 699

Db        657 YEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWI 716

Qy        700 DMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFN 759

Db        717 EVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEIN 776

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819

Db        777 NLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANS 836

Qy        820 LYL-LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN-------ISSNAIL 871

Db        837 KFIGITELKKLESKINKVFSTPIPFS---YSKN------LDCWVDNEEDIDVILKKSTIL 887

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMF 930

Db        888 NLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMF 947

Qy        931 DNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRIIWTLIDV 985

Db        948 NNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTLKDS 1007

Qy        986 NAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINS 1044

Db       1008 AGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIRE 1066

Qy       1045 SNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDT 1104

Db       1067 DNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDT 1126

Qy       1105 QYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFIIKKASNSR 1160

Db       1127 EYYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFIIKRYTPNN 1184

Qy       1161 NINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPINDDPTFYDV 1215

Db       1185 EI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP---GIPLYKK 1239

Qy       1216 LQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNYFCISQWY 1269

Db       1240 MEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDILIASNWY 1290

Qy       1270 LRRISENINKLRLGCNWQFIPVDEGWT 1296

Db       1291 FNHLKDKI----LGCDWYFVPTDEGWT 1313


RESULT 5
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             33.5%;  Score 2285;  DB 1;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 4.9e-144;
  Matches  530;  Conservative  216;  Mismatches  499;  Indels   90;  Gaps   29;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       |  ||      |||    : : |  ||:| ||    :|          ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |     |     :     :  |||:|| ||||  | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::      |  ||| ||||||   |    :    :  |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
                |:|||: ||   |  |:   :   | ::| ||   || :||:|||| ||||||| |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |     : :  | :|  | : :  |     :: |||   :|||: : ||  |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
                :|   : ::         |::   : ::           |   :: || ||| ||||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
              ::|:|: || | | |  :|     | | |:| ||    |:   |  ||         ||
Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
                | |  |           ||  :|  ::| || || |      : ||||:|:|| |
Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
              ::||||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::||||||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
              |||:||   |::|  |    || ||:|||||: :|::| | | ||: ||   : || |||
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
              ||::|| ::|  ||: ||: |||    |:::|  || ||::| : ||  |||:|:||:|
Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
              ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:||| :||  |
Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871
              |:||  |   |: :|: || |||  |   ||| || |  :  |:  |  || || : :||
Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSIL 879

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
              :| |    ||| | | : :|:|| | |: |    | :| | |:| |        :||:||::
Db        880 NLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYE 939

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKS 991
              ||| :||:| ||| |:     | ||||||:|::|:||||||:    |||| |     :
Db        940 NFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR 996

Qy        992 IFFEYSIKDNISDYINKWFSITITNDRLGNANIYIYNGSLKKSEKILNLDRINSSNDIDFK 1051
              : |:||   |||||||:|  :||||:|| |: |||||  |   : | ||  |::||:| ||
Db        997 VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFK 1056

Qy       1052 LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQ 1111
              |  | || :::||| ||:| :|||  |:  ||  ||::  ||||||: |:||  ||: |
Db       1057 LDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNL 1116

Qy       1112 GMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
                | |:   :   :  |    || :    |  :|  :|| | :|||| ::    | ||
Db       1117 YDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDN 1172

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKT 1225
              |||  | :|:|:  : :: ||:    :   ::  |    | |       |: :| |  :
Db       1173 IVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGI 1231

Qy       1226 TYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRL 1282
              |  |:: | : :    |  |  :|           :|       | || |:|    :  |
Db       1232 TNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTL 1278

Qy       1283 GCNWQFIPVDEGWTE 1297
              ||:|:|||||:|| |
Db       1279 GCSWEFIPVDDGWGE 1293


RESULT 6
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
```

```
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             33.5%;  Score 2285;  DB 2;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 4.9e-144;
  Matches  530;  Conservative  216;  Mismatches  499;  Indels   90;  Gaps   29;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       |  ||      |||    : : |  ||:| ||    :|          ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::   |:|:  ||| |     |    :     :  |||:|| ||||  | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::      |  ||| ||||||   |   :    :  |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
              |:|||: ||   |  |:   :   | ::| ||   || :||:|||| |||||| |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
              :|  : ::          |::  : ::             |   :: || ||| ||||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
              ::|:|: || | | |  :|     | | |:| ||     |:    |   ||          ||
Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
                | |  |              ||  :|  ::| || || |      : ||||:|:|| |
Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
              ::||||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::|||||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
              |||:||    |::|  |    || ||:|||||: :|::| | | ||: ||   : || |||
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
              ||::|| ::|  ||: ||: ||||    |:::|  || ||::| : ||  |||:|:||:|
Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
              ||| : :|:  |||:||| |: ||: |:||||:|||| ||| ||:|:|:|| :||  |
Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871
              |:||  |   |: :|: || ||| |   ||| || |  :  |:  |  || || : :||
Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSIL 879

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
              :| |    ||| | | : :|:|| | |: |   | :| | |:| |       :||:|::
Db        880 NLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYE 939

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKS 991
              ||| :||:| ||| |:     | ||||||:|::|:||||||:    |||||  |    :
Db        940 NFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR 996

Qy        992 IFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK 1051
              : |:||   |||||||:|  :||||:|| |: ||||| |   : | ||  |::||:|: ||
Db        997 VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFK 1056

Qy       1052 LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQ 1111
              |  | || :::||| ||:| :|||  |:  ||  ||::  ||||||: |:||  ||: |
Db       1057 LDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNL 1116

Qy       1112 GMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
                | |:   :   :  |    || :    | :|  :|| | :||||| ::    | ||
Db       1117 YDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDN 1172

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKT 1225
              |||  | :|:|:  : :: || ||:     :  ::  |   | |       |: :| |  :
Db       1173 IVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGI 1231

Qy       1226 TYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRL 1282
              |  |:: | : :    |  | :|           :|      | || |:|    :  |
Db       1232 TNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTL 1278

Qy       1283 GCNWQFIPVDEGWTE 1297
              .||:|:|||||:|| |
Db       1279 GCSWEFIPVDDGWGE 1293


RESULT 7
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-28

  Query Match             33.5%;  Score 2285;  DB 3;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 4.9e-144;
  Matches  530;  Conservative  216;  Mismatches  499;  Indels   90;  Gaps   29;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |    |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |     |  ||      |||    : : |  ||:| ||    :|           ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |      |     :     :  |||:|| ||||  | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::      | ||| ||||||   |   :    :  |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
               |:|||: ||   |  |:   :   | ::| ||   || :||:|||| |||||||
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
              :|   : ::         |::   : ::           |   :: || ||| ||||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
              ::|:|: || | | |  :|      | | |:| ||    |:   |   ||         ||
Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
                | |  |           ||  :|  ::| || || |      : ||||:|:|| |
Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
              ::||||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::||||||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
              |||:||   |::|  |    || ||:|||||: :|::|  | |  || :  : || |||
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
              ||::|| ::|  ||: ||: |||    |:::|  || ||::|  : ||  |||:|:||:|
Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
              ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:||| :||  |
Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871
              |:||  |   |: :|: || |||  |   ||| || |  :  |:  |  || || : :||
Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSIL 879

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
              :| |    ||| | | : :|:|| | |:  |   | :| | |:| |       :||:||::
Db        880 NLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYE 939

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKS 991
              ||| :||:| ||| |:     | ||||||:|::|:||||||:    |||| |     :
Db        940 NFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR 996

Qy        992 IFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK 1051
              : |:||   |||||||:|  :||||:|| |: ||||| |   : | ||  |::||:| ||
Db        997 VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFK 1056

Qy       1052 LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQ 1111
              |  | || :::||| ||:| :|||  |:  ||  ||::  ||||||: |:||  ||: |
Db       1057 LDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNL 1116

Qy       1112 GMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
                | |:   :   :  |    || :    |  :|  :|| | :|||||  ::    | ||
Db       1117 YDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDN 1172

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKT 1225
              |||  | :|:|:  : :: ||:    :   ::  |    | |       |: :| |  :
Db       1173 IVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGI 1231
```

```
Qy       1226 TYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRL 1282
Db       1232 TNKCKMNLQDNNGNDIGFIGFHQF-----------NNIAKLVASNWYNRQIER--SSRTL 1278

Qy       1283 GCNWQFIPVDEGWTE 1297
Db       1279 GCSWEFIPVDDGWGE 1293


RESULT 8
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-28

  Query Match             33.5%;  Score 2285;  DB 4;  Length 1296;
  Best Local Similarity   39.7%;  Pred. No. 4.9e-144;
  Matches  530;  Conservative  216;  Mismatches  499;  Indels   90;  Gaps   29;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
Db          6 KQFNYKDPVNGVDIAYIK-IPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871
Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSIL 879

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
Db        880 NLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYE 939

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKS 991
Db        940 NFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR 996

Qy        992 IFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK 1051
Db        997 VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFK 1056

Qy       1052 LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQ 1111
Db       1057 LDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNL 1116

Qy       1112 GMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
```

```
Db       1117 YDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDN 1172

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKT 1225

Db       1173 IVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGI 1231

Qy       1226 TYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRL 1282

Db       1232 TNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTL 1278

Qy       1283 GCNWQFIPVDEGWTE 1297

Db       1279 GCSWEFIPVDDGWGE 1293


RESULT 9
US-09-255-829-24
; Sequence 24, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-24

  Query Match           26.3%;  Score 1792.5;  DB 4;  Length 548;
  Best Local Similarity  60.8%;  Pred. No. 1.1e-111;
  Matches  336;  Conservative   98;  Mismatches  112;  Indels    7;  Gaps    4;

Qy         1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db         1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy        61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db        61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy       121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Db       121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy       181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Db       181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy       241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Db       241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy       301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359

Db       301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy       360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419

Db       361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy       420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479

Db       421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy       480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539

Db       476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy       540 DSLFEYLHAQTFP 552

Db       535 NTIFQYLYSQTFP 547


RESULT 10
US-09-255-829-14
; Sequence 14, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  14:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  953 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-14

  Query Match             22.0%;  Score 1502;  DB 4;  Length 953;
  Best Local Similarity   38.5%;  Pred. No. 6.2e-92;
  Matches  362;  Conservative  145;  Mismatches  347;  Indels   86;  Gaps   20;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       |  ||      |||    : : |  ||:| ||    :|          ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |      |    |      :     : |||:|| |||||  | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |      ||  ::       | ||| ||||||   |    :    : |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
              |:|||: ||   |  |:   :   | ::| ||   || :||:|||| |||||| |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNK------------EAYEEISLEHLVIYRIAMCKPVMYKNTGK-----SEQCIIVN 454
              :|               | |: : :  ::  :          |  |:     :: || ||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVN 465

Qy        455 NEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEP 514
              | ||||  ::|:|: || | | |  :|     | | |:| ||    |:   |   ||
Db        466 NWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISI 524

Qy        515 FTNFDDIDIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSL 564
                  ||      | |  |           ||  :|  ::| || || |       : ||||:
Db        525 ENLSSDI-----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 579

Qy        565 NDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSI 624
              |:|| | ::|||||::  |:|  |    |::|: ||: :: ||| |::: || ||::|::|
Db        580 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 639

Qy        625 IIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHII 684
              ||||||||||:||   |::|  |    || ||:|||||: :|::| | | ||:  ||   :
Db        640 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTV 699

Qy        685 MTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYN 744
              || ||| ||::|| ::|  ||: ||: |||   |:::|  || ||::| : ||  |||
Db        700 QTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYN 759

Qy        745 RYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFD 804
              :|:||:| ||| : :|:  |||:||| |: ||: |:|||:|||| |||  ||:|:|||
Db        760 QYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFD 819

Qy        805 DNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN 864
              :||  ||:||  |   |: :|: || |||  |   ||| || |  :  |:  |  || :
Db        820 ASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKS 879

Qy        865 ISSNAILSLSYRGGRLIDS--------------SGYGAT 889
                    |    |  |:|:               :|||::
```

```
Db        880 RPGPETLC----GAELVDALQFVCGDRGFYFNKPTGYGSS 915


RESULT 11
US-09-255-829-8
; Sequence 8, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-8

  Query Match             22.0%;  Score 1499;  DB 4;  Length 871;
  Best Local Similarity   39.6%;  Pred. No. 8.6e-92;
  Matches  355;  Conservative  140;  Mismatches  341;  Indels   60;  Gaps   18;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||   | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: :
Db          1 MPFVNKQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGD 58

Qy         61 ASTGVFSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
                 :    :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||:
Db         59 LNPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFW 118

Qy        120 GNAS--TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM 177
              | ::  |       |  ||      |||    : : |  ||:| ||    :|
Db        119 GGSTIDTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKS 167

Qy        178 NGHSPIS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
              ||  ::    |:|:   ||| |      |    |      :     : |||:|| |||||
Db        168 FGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIH 227

Qy        235 VLHGLYGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIY 289
              | |||| |: |      ||  ::       |  ||| ||||||   |    :    : |
Db        228 AGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYY 287
```

```
Qy        290 NKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYK 349
              ||    |:|||: ||   |  |:   :   | ::| ||   || :||:|||| |||||||
Db        288 NK----FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYK 343

Qy        350 ALMFGFTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTE 407
              |   :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||
Db        344 MLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAAN 400

Qy        408 FNGQNKAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNE 456
              |||||  :|   : ::         |::   : ::           |   :: || |||
Db        401 FNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNW 460

Qy        457 DLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFT 516
              ||||  ::|:|: || | | |  :|     | | |:| ||    |:   |   ||
Db        461 DLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIEN 519

Qy        517 NFDDIDIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLND 566
              ||      | |  |            ||  :|  ::| || || |      : ||||:|:
Db        520 LSSDI-----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNE 574

Qy        567 ALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIII 626
              || | ::|||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::|||
Db        575 ALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIII 634

Qy        627 PYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMT 686
              ||||||||:||   |::|  |    || ||:|||||: :|::|  | | ||: ||  : |
Db        635 PYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQT 694

Qy        687 ISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRY 746
              | ||| ||::|| ::|  ||: ||: |||| |:::|  || ||::|  : ||  |||:|
Db        695 IDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQY 754

Qy        747 SEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDN 806
              :||:| ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:||| :
Db        755 TEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        807 LKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYI 862
              ||  ||:||   |   |: :|: || |||  |   ||| || |  :  |:  |  ||
Db        815 LKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYI 870
```

RESULT 12
US-09-255-829-18
; Sequence 18, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1013 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-18

Query Match 22.0%; Score 1497.5; DB 4; Length 1013;
Best Local Similarity 36.0%; Pred. No. 1.3e-91;
Matches 378; Conservative 169; Mismatches 395; Indels 109; Gaps 24;

```
Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |      |  ||      |||   : : |  ||:| ||    :|            ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |      |     |      :     : |||:|| ||||  | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |      ||  ::       |  ||| ||||||   |    :   :  |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
               |:|||: ||   |  |:   :   | ::| ||   || :||:|||| ||||||| |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNK------------EAYEEISLEHLVIYRIAMCKPVMYKNTGK-----SEQCIIVN 454
              :|               | |: : :  ::  :          |  |:     :: || ||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVN 465

Qy        455 NEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEP 514
              | ||||  ::|:|: || | | |  :|     | | |:| ||    |:   |   ||
Db        466 NWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISI 524

Qy        515 FTNFDDIDIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSL 564
                   ||      | |  |           ||  :|  ::| || || |      : ||||:
Db        525 ENLSSDI-----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 579

Qy        565 NDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSI 624
              |:|| | ::||||||:: |:| |     |::|: ||: :: ||| |::: || ||::|::|
Db        580 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 639

Qy        625 IIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHII 684
              ||||||||||:||   |::|  |    || ||:|||||: :|::| | | ||: ||   :
Db        640 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTV 699

Qy        685 MTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYN 744
              || ||| ||::|| ::|  ||: ||: ||||   |:::|  || ||::|  : ||  |||
Db        700 QTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYN 759

Qy        745 RYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFD 804
              :|:||:| ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:|||
Db        760 QYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFD 819
```

```
Qy        805 DNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN 864
              :||  ||:||  |   |: :|: || |||  |  ||| || |  :  |:  |  || :
Db        820 ASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKS 879

Qy        865 ISSNAILSLSYRGGRLIDSSGYGATMNVGSDVI----FNDIGNGQF----KLNNSENSNI 916
                ::   :   :   :|:       |   :::     |:     |    | : |:::|:
Db        880 GLNSPGAAHYAQHDEAVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANL 939

Qy        917 TAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGN 976
              |   |  : |:           : || :|   :      | |:
Db        940 LAEAKK--LNDA-----------QAPKVDNKFNKEQQNAFYEILH--------------- 971

Qy        977 RIIWTLIDVNAKSKSIFFEYSIKDNISDYIN 1007
                  | ::| : :: | : |:||: |   |
Db        972 -----LPNLNEEQRNAFIQ-SLKDDPSQSAN 996


RESULT 13
US-09-255-829-16
; Sequence 16, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-16

  Query Match             21.9%;  Score 1494.5;  DB 4;  Length 907;
  Best Local Similarity   39.4%;  Pred. No. 1.8e-91;
  Matches  355;  Conservative  140;  Mismatches  338;  Indels   67;  Gaps   18;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
                :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       | ||      |||    : : |  ||:| ||    :|          ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |     |     |      :      :  |||:|| ||||  | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::      | ||| ||||||   |    :    : |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
              |:|||: ||   | :|:  :   | ::| ||   || :||:|||| |||||||  |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |   :  :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNK------------EAYEEISLEHLVIYRIAMCKPVMYKNTGK-----SEQCIIVN 454
                :|              | |: : :  ::  :          |  |:     :: || ||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKIEGRCDGALNDLCIKVN 465

Qy        455 NEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEP 514
              | ||||  ::|:|: || | | |  :|      | | |:| ||    |:   |    ||
Db        466 NWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISI 524

Qy        515 FTNFDDIDIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSL 564
                ||    | | |              ||  :|  ::| || || |       : ||||:
Db        525 ENLSSDI-----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSV 579

Qy        565 NDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSI 624
              |:|| | ::||||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::|
Db        580 NEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITI 639

Qy        625 IIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHII 684
              ||||||||||:||   |::|  |    || ||:|||||: :|::| | | ||: ||   :
Db        640 IIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTV 699

Qy        685 MTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYN 744
              || ||| ||::|| ::|  ||: ||: ||||    |:::|  || ||::| : ||  |||
Db        700 QTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYN 759

Qy        745 RYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFD 804
              :|:||:| ||| : :|:  |||:||| |: ||: |:||||:||||| ||| ||:|:|||
Db        760 QYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFD 819

Qy        805 DNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN 864
              :||  ||:||  |   |: :|: || |||  |  ||| || |  :  |:  |  || :
Db        820 ASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKS 879


RESULT 14
US-09-255-829-2
; Sequence 2, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-2

  Query Match             21.9%;  Score 1494;  DB 4;  Length 871;
  Best Local Similarity   39.6%;  Pred. No. 1.9e-91;
  Matches  353;  Conservative  140;  Mismatches  338;  Indels   60;  Gaps   18;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::  . : |     |||:| ::||::||| |:   |:: : :
Db          6 KQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       |  ||      |||    : : |  ||:| ||    :|             ||
Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |     |     |     :     : |||:|| ||||| | |
Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::       |  ||| ||||||   |    :    : |||
Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
              |:|||: ||   |  |:   :   | ::| ||   || :||:|||| |||||||  |
Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |     :: |||   :|||: : ||   |||||
Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
              :|   : ::        |::   : ::              |   :: || ||| ||||
Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
              ::|:|: || | | |  :|     | | |:| ||    |:  |   ||        ||
Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
                  | |  |           ||  :|  ::| || || |      : ||||:|:|| |
Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
              ::||||||:: |:| |    |::|: ||: :: ||| |::: || ||::|::|||||||
Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
              |||:||   |::|  |    || ||:|||||: :|::| | | ||: ||   : || |||
Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
              ||::|| ::|  ||: ||: ||||   |:::|  || ||::| : ||  |||:|:||:|
Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
              ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:||| :||  |
Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYI 862
              |:||  |   |: :|: || |||  |   ||| || |  :  |:  |  ||
Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYI 870


RESULT 15
US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
;  GENERAL INFORMATION:
;    APPLICANT:  Shone, Clifford Charles
;    APPLICANT:  Quinn, Conrad Padraig
;    APPLICANT:  Foster, Keith Alan
;    TITLE OF INVENTION:  Recombinant Toxin Fragments
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/255,829
;      FILING DATE:  23-FEB-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB97/02273
;      FILING DATE:  22-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,893
;      FILING DATE:  27-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1581.0130002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-255-829-26

  Query Match             21.9%;  Score 1494;  DB 4;  Length 871;
  Best Local Similarity   39.6%;  Pred. No. 1.9e-91;
  Matches  353;  Conservative  140;  Mismatches  338;  Indels   60;  Gaps   18;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
```

```
Db          6 KQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 63

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123

Db         64 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 123

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182

Db        124 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 172

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239

Db        173 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 232

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294

Db        233 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 289

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354

Db        290 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 348

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412

Db        349 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 405

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461

Db        406 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 465

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521

Db        466 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 524

Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571

Db        525 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 579

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631

Db        580 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 639

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691

Db        640 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 699

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751

Db        700 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 759

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811

Db        760 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 819

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYI 862

Db        820 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYI 870


Search completed: September 10, 2004, 03:47:16
Job time : 31.5291 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 251.059 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 6816
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6816	100.0	1297	12	US-10-272-898-77	Sequence 77, Appl
2	6816	100.0	1297	15	US-10-354-774-77	Sequence 77, Appl
3	6816	100.0	1297	15	US-10-271-012-77	Sequence 77, Appl
4	6816	100.0	1297	16	US-10-729-122-77	Sequence 77, Appl
5	6816	100.0	1297	16	US-10-729-039-77	Sequence 77, Appl
6	6816	100.0	2017	15	US-10-452-024-146	Sequence 146, App
7	6809	99.9	1297	15	US-10-452-024-7	Sequence 7, Appli
8	6809	99.9	1297	15	US-10-452-024-150	Sequence 150, App
9	6809	99.9	1297	15	US-10-205-516-14	Sequence 14, Appl
10	6797	99.7	1307	15	US-10-205-516-28	Sequence 28, Appl
11	4409.5	64.7	848	13	US-10-011-588-45	Sequence 45, Appl
12	4007	58.8	1291	15	US-10-452-024-123	Sequence 123, App
13	4006	58.8	1291	15	US-10-452-024-121	Sequence 121, App
14	4001	58.7	1291	15	US-10-452-024-142	Sequence 142, App
15	3997	58.6	1291	15	US-10-452-024-2	Sequence 2, Appli
16	3997	58.6	1291	15	US-10-452-024-118	Sequence 118, App
17	3996	58.6	1291	15	US-10-452-024-122	Sequence 122, App
18	3992	58.6	1290	15	US-10-360-101-220	Sequence 220, App
19	3992	58.6	1290	15	US-10-452-024-119	Sequence 119, App
20	3987	58.5	1291	12	US-10-272-898-42	Sequence 42, Appl
21	3987	58.5	1291	15	US-10-354-774-42	Sequence 42, Appl
22	3987	58.5	1291	15	US-10-271-012-42	Sequence 42, Appl
23	3987	58.5	1291	16	US-10-729-122-42	Sequence 42, Appl
24	3987	58.5	1291	16	US-10-729-039-42	Sequence 42, Appl
25	3932	57.7	1291	12	US-10-272-898-40	Sequence 40, Appl
26	3932	57.7	1291	15	US-10-354-774-40	Sequence 40, Appl
27	3932	57.7	1291	15	US-10-271-012-40	Sequence 40, Appl
28	3932	57.7	1291	15	US-10-452-024-117	Sequence 117, App
29	3932	57.7	1291	15	US-10-205-516-4	Sequence 4, Appli
30	3932	57.7	1291	16	US-10-729-122-40	Sequence 40, Appl
31	3932	57.7	1291	16	US-10-729-039-40	Sequence 40, Appl
32	3931	57.7	1301	15	US-10-205-516-18	Sequence 18, Appl
33	3684.5	54.1	1169	14	US-10-241-596-20	Sequence 20, Appl
34	2839.5	41.7	876	14	US-10-241-596-106	Sequence 106, App
35	2839.5	41.7	876	14	US-10-241-596-108	Sequence 108, App
36	2838.5	41.6	876	14	US-10-241-596-82	Sequence 82, Appl
37	2836.5	41.6	858	14	US-10-241-596-22	Sequence 22, Appl
38	2834.5	41.6	864	14	US-10-241-596-102	Sequence 102, App
39	2834	41.6	888	14	US-10-241-596-112	Sequence 112, App
40	2833.5	41.6	860	14	US-10-241-596-175	Sequence 175, App
41	2833.5	41.6	862	14	US-10-241-596-94	Sequence 94, Appl
42	2833.5	41.6	866	14	US-10-241-596-88	Sequence 88, Appl
43	2833.5	41.6	867	14	US-10-241-596-96	Sequence 96, Appl
44	2833.5	41.6	867	14	US-10-241-596-98	Sequence 98, Appl
45	2833.5	41.6	870	14	US-10-241-596-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-77
; Sequence 77, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-272-898-77

  Query Match             100.0%;  Score 6816;  DB 12;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 2
US-10-354-774-77
; Sequence 77, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-354-774-77

  Query Match             100.0%;  Score 6816;  DB 15;  Length 1297;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 3
US-10-271-012-77
; Sequence 77, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-271-012-77

```
  Query Match           100.0%;  Score 6816;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
```

```
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 4
US-10-729-122-77
; Sequence 77, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-122-77

Query Match 100.0%; Score 6816; DB 16; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
```

```
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

```
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
```

```
RESULT 5
US-10-729-039-77
; Sequence 77, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-039-77

  Query Match           100.0%;  Score 6816;  DB 16;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
```

```
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 6
US-10-452-024-146
; Sequence 146, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 146
;   LENGTH: 2017
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-146

  Query Match             100.0%;  Score 6816;  DB 15;  Length 2017;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1297;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 780

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 840

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 900

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 960

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 1020

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 1080

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 1140

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 1200

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 1260

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 1320

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 1380

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 1440

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 1500
```

```
Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 1560

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 1620

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 1680

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1740

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1800

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1860

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1920

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1980

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1981 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 2017
```

RESULT 7
US-10-452-024-7
; Sequence 7, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1297
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-7

Query Match 99.9%; Score 6809; DB 15; Length 1297;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
```

```
Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
             |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 8
US-10-452-024-150
; Sequence 150, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 150
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-150

  Query Match             99.9%;  Score 6809;  DB 15;  Length 1297;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1296;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 9
US-10-205-516-14
; Sequence 14, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
```

```
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7)
;   OTHER INFORMATION: mixed sequence
US-10-205-516-14

  Query Match             99.9%;  Score 6809;  DB 15;  Length 1297;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1296;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 10
US-10-205-516-28
; Sequence 28, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
;   LENGTH: 1307
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7)
;   OTHER INFORMATION: mixed sequence
US-10-205-516-28

  Query Match             99.7%;  Score 6797;  DB 15;  Length 1307;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1296;  Conservative    0;  Mismatches    1;  Indels    4;  Gaps    1;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
```

```
Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYK----NTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAET 476
              ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKIEGRNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAET 480

Qy        477 IAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIF 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 IAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIF 540

Qy        537 VDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFV 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFV 600

Qy        597 NWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAIL 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 NWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAIL 660

Qy        657 MEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQ 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 MEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQ 720

Qy        717 FYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNI 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 FYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNI 780

Qy        777 DDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRH 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 DDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRH 840

Qy        837 LKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDV 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDV 900

Qy        897 IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNE 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNE 960

Qy        957 YTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITN 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 YTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITN 1020

Qy       1017 DRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNA 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 DRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNA 1080

Qy       1077 TEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFN 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFN 1140

Qy       1137 NAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKE 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKE 1200

Qy       1197 IQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVW 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 IQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVW 1260

Qy       1257 DTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||||||
Db       1261 DTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1301


RESULT 11
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
;  APPLICANT: Smith, Leonard
;  APPLICANT:  Jensen, Melody
;  TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
;  TITLE OF INVENTION:  RESEARCH AND CLINICAL THERAPY
;  FILE REFERENCE: A34796 067252.0113
;  CURRENT APPLICATION NUMBER: US/10/011,588
;  CURRENT FILING DATE:  2002-03-29
;  PRIOR APPLICATION NUMBER: 09/910,186
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/246,744
;  PRIOR FILING DATE: 2000-11-06
;  PRIOR APPLICATION NUMBER: 60/311,966
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 848
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45

  Query Match           64.7%;  Score 4409.5;  DB 13;  Length 848;
  Best Local Similarity  99.9%;  Pred. No. 1.6e-303;
  Matches  848;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIEN-QLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 599

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        600 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 659

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Db        660 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 719

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNININIDFNDIDFKLNQSINLAINNIDDFI 780

Db        720 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNININIDFNDIDFKLNQSINLAINNIDDFI 779

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Db        780 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 839

Qy        841 IPFDLSLYT 849

Db        840 IPFDLSLYT 848


RESULT 12
US-10-452-024-123
; Sequence 123, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 123
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-123

  Query Match           58.8%;  Score 4007;  DB 15;  Length 1291;
  Best Local Similarity  58.0%;  Pred. No. 1e-274;
  Matches  756;  Conservative  249;  Mismatches  280;  Indels   18;  Gaps   11;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Db        241 GIKVNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359

Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419

Db        361 IAENYKIKTRAAYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479

Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIAY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539

Db        476 NTQNNYIENDFSINELILDTDLISKIELPSENTESLTDF-NVYVPVYKKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599

Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659

Db        595 KQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719

Db        655 IPELLIPVVGAFLLESYIDNKNKIIETINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNININIDFNDIDFKLNQSINLAINNIDDF 779

Db        715 IKEGMYKALNYQAQALEEIIKYKYNIYSEKERSNININIDFNDVNSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839

Db        775 INECSVSYLMKKMIPLAVEKLLDFDNTLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899

Db        835 SIPFDLSTYTNNTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959

Db        895 D--KNQFKLTSSANSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019

Db        953 INCMKNNSGWKISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS- 1011

Qy       1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079

Db       1012 DNAKIYINGKLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139

Db       1072 EEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197

Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYMYKYFKKE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254

Db       1190 EEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1249 VFKEYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 13
US-10-452-024-121
; Sequence 121, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
```

```
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 121
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-121

  Query Match             58.8%;  Score 4006;  DB 15;  Length 1291;
  Best Local Similarity   58.0%;  Pred. No. 1.2e-274;
  Matches  756;  Conservative  249;  Mismatches  280;  Indels   18;  Gaps   11;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | |||||||||:|::|||||||    | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::||:||:: |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::| |:|:|::|| |  || :: ||:  |||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                  ||||  | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||:::||| || |:|||| :| :|||||||||| |||:|||||| :||:| ||||: |
Db        241 GIKVNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:|||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| ||||: |||:| : |||| |||  |||||: ||:: |: |||||:||:|
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  |  |    || |:|||||||:|:||| ||:| | |||
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIAY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              ||||| |||:|||::|||| || | |:||:|||| |:| :: :||| || |:|||| |
Db        476 NTQNNYIENDFSINELILDTDLISKIELPSENTESLTDF-NVYVPVYKKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | :::||  |: : ||:||::|:|:|:|||| |||||||||| ||:||||| ||:||:||
Db        595 KQIVNDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | |||||: ||  || ||::|| |||:|| |||||||:|||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIETINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::: ||||||||:: |||: || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYKYNIYSEKERSNINIDFNDVNSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              ||:||:|||| :|||||:|| |||: |:::|| ||| |:|||:      ||||:::||
Db        775 INECSVSYLMKKMIPLAVEKLLDFDNTLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
              ||||||| || :||||::|| | |:| :| ||:| ||  :||| ||||| : |    |  |
Db        835 SIPFDLSTYTNNTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
              |    |||| :| || |    |:: ::::||| :||::||:| ||| |: || |: |||||
Db        895 D--KNQFKLTSSANSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
              |:|:||:||||:||:|| |||||||:| | ||:|||||||::||:|||:|| :||||:
Db        953 INCMKNNSGWKISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS- 1011

Qy       1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
              || ||||| |:    | ::  : ::::| |||    | |:|:|:| |:||  ||: : :
Db       1012 DNAKIYINGKLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
                :| ||| :  |||||||||| |: :||:|| | :| |||    :|:||   |: :|  :
Db       1072 EEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
                |||::||:| :|||:: |||::| ||:|||: |||||:  |:: | :|||:     |:
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYMYKYFKKE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
                : :|||||:|    ||: :|||:| |: ||:||:| :||   |    || || :| :  |
Db       1190 EEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
                |:  | :||||:|||: :         :|||||||| ||||||
Db       1249 VFKEYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 14
US-10-452-024-142
; Sequence 142, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 142
;   LENGTH: 1291
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-452-024-142

  Query Match             58.7%;  Score 4001;  DB 15;  Length 1291;
  Best Local Similarity   58.0%;  Pred. No. 2.7e-274;
  Matches  756;  Conservative  235;  Mismatches  294;  Indels   18;  Gaps   10;

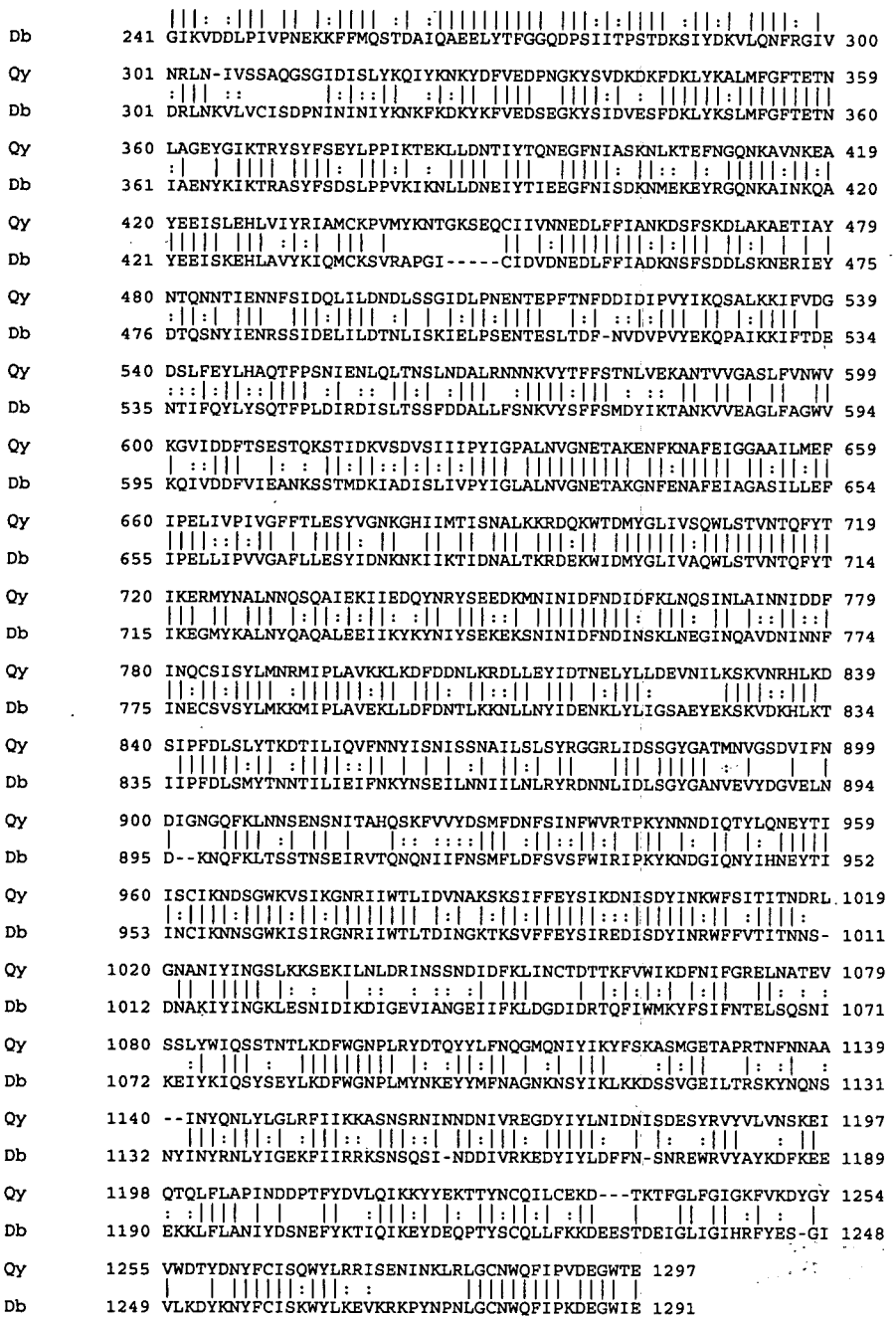
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | |||||||||:|::|||||||    | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::||:||:: |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::| |:|:|::|| |  || :: ||:  |||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                  ||||  | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

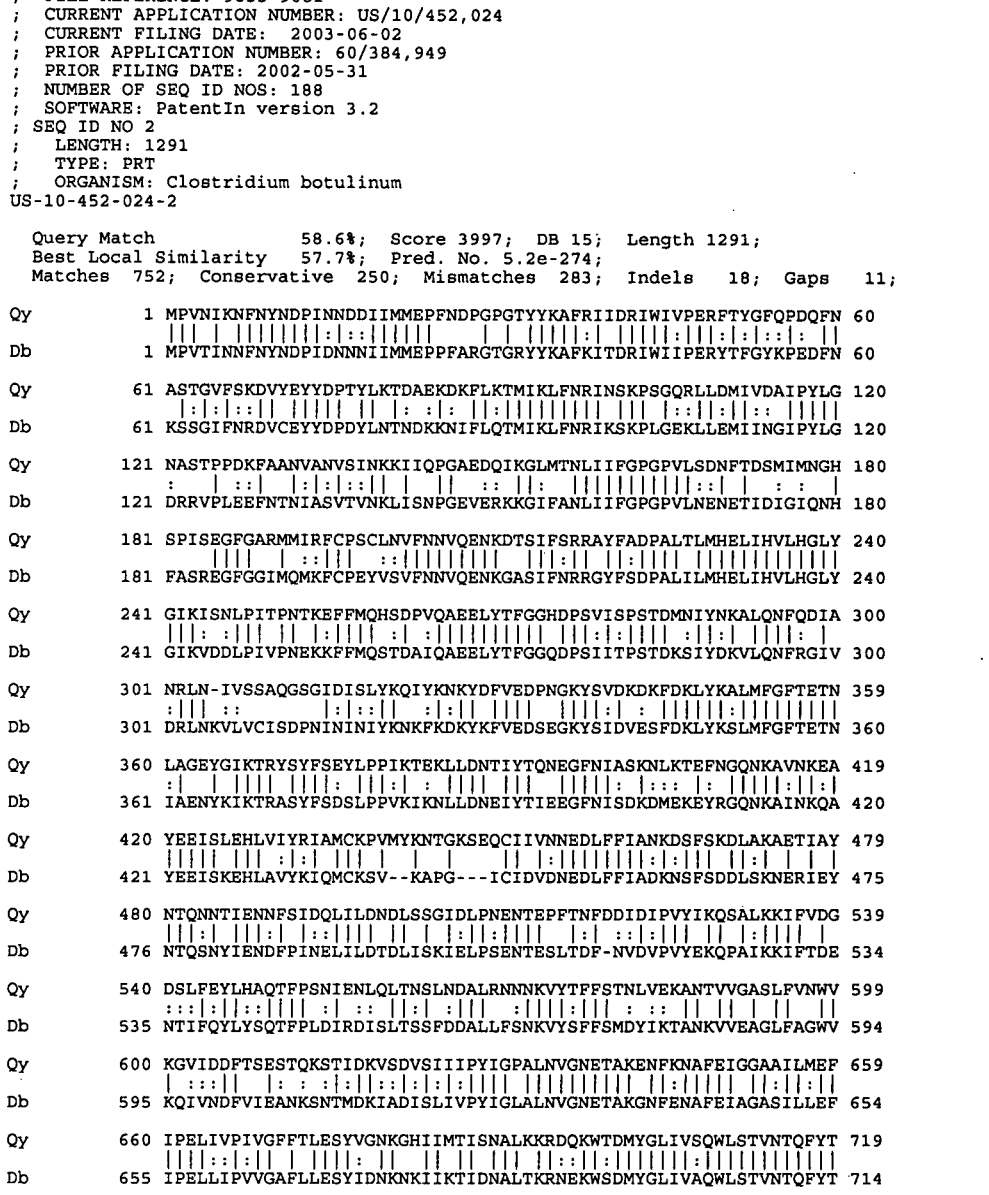
Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
```



RESULT 15

US-10-452-024-2
; Sequence 2, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-2

Query Match 58.6%; Score 3997; DB 15; Length 1291;
Best Local Similarity 57.7%; Pred. No. 5.2e-274;
Matches 752; Conservative 250; Mismatches 283; Indels 18; Gaps 11;



```
Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::| |||||||||: |||: || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              || ||:|||| :||||||:|| |||: ||::|| ||| |:|||:       |||||::||
Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
               :|||||:|| |||||::|| | | | :| ||:| |:   ||| ||||| : |    |  |
Db        835 IMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
              |    |||| :| || |   |:: ::::|:| :||::||:| ||| |: || |: |||||
Db        895 D--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
              |:|:||:||||:||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: |
Db        953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy       1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
               || ||||| |: :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :
Db       1012 NNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
                | ||| :  |||||||| |: :||:|| | :| |||    : :||   |: :|  :
Db       1072 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
                |||::||:| :|||:: |||::| ||:|||: |||||:  |:: | :|||      |:
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
               : :||||||:|   ||: :|||:| |: ||:||:| :||   |   || || :| :  |
Db       1190 EEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |:: | :|||||:|||: :      |:|||||||| ||||||
Db       1249 VFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

Search completed: September 10, 2004, 04:58:22
Job time : 255.059 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 22.9134 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 6816
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6816	100.0	1297	2	S39791	neurotoxin - Clost
2	3997	58.6	1291	1	A48940	bontoxilysin (EC 3
3	3932	57.7	1291	2	I40631	non-proteolytic bo
4	2540.5	37.3	1315	1	BTCLTN	tentoxilysin (EC 3
5	2328.5	34.2	1268	2	S33411	botulinum neurotox
6	2291	33.6	1296	1	BTCLAB	bontoxilysin (EC 3
7	2257	33.1	1252	2	S21178	botulinum neurotox
8	2233.5	32.8	1251	2	JH0256	botulinum neurotox
9	2211	32.4	1296	2	I40645	botulinum neurotox
10	2188.5	32.1	1274	2	I40813	neurotoxin type F
11	2081.5	30.5	1276	2	S11455	botulinum neurotox
12	2028	29.8	1285	2	S70582	botulinum neurotox
13	1908	28.0	1291	2	S46431	botulinum neurotox
14	1908	28.0	1291	2	A49777	botulinum neurotox
15	791.5	11.6	367	2	S48106	neurotoxin type E
16	789.5	11.6	366	2	S48110	neurotoxin type F
17	738	10.8	369	2	S48109	neurotoxin type F
18	692.5	10.2	552	2	A49928	neurotoxin C1, BoN
19	680.5	10.0	1162	2	I40817	botulinum toxin no
20	673.5	9.9	1162	2	A47708	progenitor toxin n
21	640	9.4	1193	2	S68218	botulinum neurotox
22	625	9.2	1165	2	I40644	botulinum neurotox
23	619	9.1	1196	2	JQ1467	toxin, nontoxic co
24	618	9.1	1196	2	S46430	botulinum neurotox
25	602	8.8	1193	2	JC4901	nontoxic-nonhemagg
26	333	4.9	1127	2	T28317	ORF MSV156 hypothe
27	304	4.5	2485	1	H71621	serine/threonine-s
28	293.5	4.3	4688	2	F82885	hypothetical prote
29	290.5	4.3	2401	2	T28676	rhoptry protein -
30	281	4.1	2819	2	T09080	probable chloroqui
31	280.5	4.1	3394	2	T18501	hypothetical prote
32	278	4.1	2539	2	B71619	hypothetical prote
33	276	4.0	2178	2	S55805	alpha-toxin - Clos
34	274	4.0	2599	2	F90608	ABC transporter pe
35	273	4.0	2708	2	T09079	probable chloroqui
36	272.5	4.0	2710	2	A37052	toxin A - Clostrid
37	269.5	4.0	2269	2	T28677	rhoptry protein -
38	267	3.9	1830	2	E82909	conserved hypothet
39	264	3.9	4981	2	T18489	hypothetical prote
40	263.5	3.9	971	2	A70179	exodeoxyribonuclea
41	263	3.9	2013	2	C71610	probable membrane
42	263	3.9	2819	2	A90551	conserved hypothet
43	260	3.8	2136	2	A05037	hypothetical prote
44	257	3.8	1436	2	D71618	hypothetical prote
45	256.5	3.8	1817	2	H71611	probable secreted

ALIGNMENTS

RESULT 1
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 100.0%; Score 6816; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 8.9e-299;
Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Qy        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540

Qy        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720

Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 1140

Qy       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 1200

Qy       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 1260

Qy       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |||||||||||||||||||||||||||||||||||||
Db       1261 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 2
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific ide
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisior
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus tox
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with t
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synaps
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
```

F;231/Active site: Glu #status predicted

Query Match 58.6%; Score 3997; DB 1; Length 1291;
Best Local Similarity 57.7%; Pred. No. 4.4e-172;
Matches 752; Conservative 250; Mismatches 283; Indels 18; Gaps 11;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | ||||||||:|::||||||    | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
               |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::||:||:: |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::|  |:|:|::|| |  ||  :: ||:  |||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                |||| | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||: :||| || |:|||| :| :||||||||| |||:|:|||| :||:| ||||: |
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||  ||||:| : ||||||:|||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| |||||: |||:| : |||| |||  |||||: |::: |: |||||:||:|
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  |  |    || |:||||||||:|:||| ||:| | | |
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              |||:| |||:| |::|||| || | |:||:|||  |:| ::|:||| || |:|||| |
Db        476 NTQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | :::||  |: : :|:||::|:|:|:|||| |||||||||| ||:||||| ||:||:||
Db        595 KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | |||||:  ||  || || ||| ||::||:|||||||:||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::| |||||||||: ||| || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              || ||:|||| :||||||:|| |||: ||::|| ||| |:|||:       ||||::||
Db        775 INGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
               :|||||:|| ||||::|| | | | :| ||:| |:   ||| ||||| : |   | |
Db        835 IMPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
              |    |||| :| || |   |:: ::::|:| :||::||:| ||| |: || |: |||||
Db        895 D--KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
              |:|:||:||||:||:||||||||||:| |:||:|||||:|:::||:|||:|| :||||: |
Db        953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy       1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
              || ||||| |: :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :
Db       1012 NNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
                 | ||| :  |||||||| |: :||:|| | :| |||    : :||   |: :|  :
Db       1072 EERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
                 |||::||:| :|||:: |||::| ||:|||: |||||:   |:: | :|||     |:
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
              : :||||||:|   ||: :|||:| |: ||:||:| :||   |   || || :| :  |
Db       1190 EEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |:: | :|||||:|||: :      |:|||||||| ||||||
Db       1249 VFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291
```

RESULT 3
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 57.7%; Score 3932; DB 2; Length 1291;
Best Local Similarity 56.9%; Pred. No. 3.7e-169;
Matches 741; Conservative 247; Mismatches 297; Indels 18; Gaps 10;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | ||||||||:||:||||||    | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
```

```
                |:|:|::|| ||||| || |: :|: | :|:|||||| ||| |::||:||:: |||||
Db        61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFFQTLIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy       121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
             :   | ::|  |:|:|::|| |  ||  :: ||:  |||||||||||::| |  : :  |
Db       121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy       181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                 |||| | ::||| ::|||||||||| |||:|| ||:|||| ||||||||||||
Db       181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy       241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
             |||: :||| || |:|||| :| :|||||||||| |||:|||||| :||:| |||| : |
Db       241 GIKVDDLPIVPNEKKFFMQSTDTIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIV 300

Qy       301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
             :||| ::       |:|::||  :|:|| |||| ||||:| : |:|||||:|| |||| |
Db       301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFNKLYKSLMLGFTEIN 360

Qy       360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
             :|  | |||| |||: |||:| : |||| |||  |||||: ||:  |: |||||:||:|
Db       361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMGKEYRGQNKAINKQA 420

Qy       420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
             ||||| ||| :|:| ||| |           || |:||:|||||:|:||| ||:| | : |
Db       421 YEEISKEHLAVYKIQMCKSVKVPGI-----CIDVDNENLFFIADKNSFSDDLSKNERVEY 475

Qy       480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
             ||||| | |:| |::|||| || | |:||:|||| |:| ::|:||| || |:||:| |
Db       476 NTQNNYIGNDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKVFTDE 534

Qy       540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
             :::|:|||::|||| || :: ||:| :|||  ::|||:||| : :: || || | ||  ||
Db       535 NTIFQYLYSQTFPLNIRDISLTSSFDDALLVSSKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy       600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
             | ::|||  |: : ||:||::|:|:|:|||| |||||:|||| ||::|||| |::||:||
Db       595 KQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGDETAKGNFESAFEIAGSSILLEF 654

Qy       660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
             ||||::|:|| | ||||: ||  || || ||| || :|| |||||||:||||||||||
Db       655 IPELLIPVVGVFLLESYIDNKNKIIKTIDNALTKRVEKWIDMYGLIVAQWLSTVNTQFYT 714

Qy       720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
             ||| || ||| |:||:|:||: :|| ||||:| ||||:|||||: |||  || |::||:||
Db       715 IKEGMYKALNYQAQALEEIIKYKYNIYSEEEKSNININFNDINSKLNDGINQAMDNINDF 774

Qy       780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
             ||:||:|||| :||||||||| |||: ||::|| ||| |:|||:  |   ||||:::||
Db       775 INECSVSYLMKKMIPLAVKKLLDFDNTLKKNLLNYIDENKLYLIGSVEDEKSKVDKYLKT 834

Qy       840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
             |||||| |:   |||:|| | | | :| ||:| ||   ||| ||||| : |   | |
Db       835 IIPFDLSTYSNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN 894

Qy       900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
             |    |||| :| :| |   |:: ::::||| :||::||:| ||| |:||| |: |||||
Db       895 D--KNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTI 952

Qy       960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
             |:|:||:||||:||:|||||||||||:| |:||:||||:|:::||:|||:|| :|||| |
Db       953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy      1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
             || |||||:|: :  | ::  :  : :| |||     | |:|:|:| |:||  :|| : :
Db      1012 DNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNI 1071

Qy      1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
               :| ||| :  |||||||||| |: :||:|| | :| |||    :|:||   |: :|  :
Db      1072 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNS 1131

Qy      1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
                |||:|||:| :|||:: |||::| ||:|||: |||:|::  :  | :|||     ||

Db      1132 NYINYRNLYIGEKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQ 1189

Qy      1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
               : :|||: |:|   ||  ::||:| |: :|:||:| :||   |   || || :| :  |
Db      1190 EEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GV 1248

Qy      1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
               :   | :|||||:|||: :     |  |||||||| ||||||
Db      1249 LRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 4
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botul
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Es
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal h
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
```

```
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglic
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match           37.3%;  Score 2540.5;  DB 1;  Length 1315;
  Best Local Similarity  40.1%;  Pred. No. 1.3e-106;
  Matches  540;  Conservative  253;  Mismatches  469;  Indels   85;  Gaps   25;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||: | || |:||:||| ||||||    |  |||||:| |||||||||: :| :|: ||
Db          1 MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
               : :  :  |||| ||:||::||:||:||:|||||| :  :|: ||| |::||||||
Db         61 PPSSLI-EGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLG 119

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM--- 177
              |: :  |||  |  :|| |       ::|||||||||||: |    :::
Db        120 NSYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVD 179

Qy        178 -NGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVL 236
               : |  :|||: | : |||  :  |:|| ||  :    :  || |||| |||||||||
Db        180 NKNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVL 239

Qy        237 HGLYGIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNF 296
              |||||:::|:  | |: :| :|||: |: ||||:|||| | ::||     ::| | | ::
Db        240 HGLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDY 299

Qy        297 QDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFT 356
              : |||:|: |:|    |||  |||||: || | :| ||:|  |::|||  || ::|:|||
Db        300 KAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFT 359

Qy        357 ETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVN 416
              |  |  :: |||| ||||     |:|   |||:|||   ||||| ||:||:|:  ||| ||
Db        360 EIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVN 419

Qy        417 KEAYEEISLEHLVIYRIAMCKPV---------MYKNTGK-----SEQCIIVNNEDLFFIA 462
              |:  :    ||   | :|| :          :|  |         | || : |||| |||
Db        420 TNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFIA 479

Qy        463 NKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDID 522
              |:|||::  :  | ::|||:|  :  |:|:|::|:| :|  | | |||: | | |    |
Db        480 EKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTK--GIP 537

Qy        523 IPVYIKQSALKKI---FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFS 579
                | :|   |    :| :::::||:||  |: :: : :|||::||| |: |:|::|
Db        538 YAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYF- 596

Qy        580 TNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNET 639
               ::: | |       ||: ||: :|||||:||:||:||||:|||| |:||||||||:  :
Db        597 PSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQG 656

Qy        640 AKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWT 699
              : ||  | |  |  :|:|:|||: :|::   ::      |  || || | |:|| :||
Db        657 YEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWI 716

Qy        700 DMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFN 759
              ::| |: ::|| ||||||     :|| :|  |  ||:|||: :|  ||  ||  |  |  |
Db        717 EVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEIN 776

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819
              ::  || :  | |: ||: |: : | |:|:|:||  | |:| :||    |  |::||  |
Db        777 NLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANS 836

Qy        820 LYL-LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN-------ISSNAIL 871
               :: : |:  |:||:|:      |||    |:|:       : :: |       :  : ||
Db        837 KFIGITELKKLESKINKVFSTPIPFS---YSKN------LDCWVDNEEDIDVILKKSTIL 887

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMF 930
              :|      : | ||: :::    |  :   |       | |:|:| :  |::  : |: ||
Db        888 NLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMF 947

Qy        931 DNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRIIWTLIDV 985
              :||:::||:|  || : : ::  |  |||:||| :|       ||| ||:||| :|||| |
Db        948 NNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTLKDS 1007

Qy        986 NAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINS 1044
                : : | |   : |  : |: |||  |||||||| :||:|||| |  | :|  |  |
Db       1008 AGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIRE 1066

Qy       1045 SNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDT 1104
               |:|  ||  | :  ::| |  | || : ||  |:  ||     |   |:|||||||||||
Db       1067 DNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDT 1126

Qy       1105 QYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFIIKKASNSR 1160
              :|||       ::: :|  :        ||  :: |  :|  |: || ||:||||: : :
Db       1127 EYYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFIIKRYTPNN 1184

Qy       1161 NINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPINDDPTFYDV 1215
               |  |: |: ||:| | :   ::|    |       |   |    : ||       |
Db       1185 EI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP---GIPLYKK 1239

Qy       1216 LQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNYFCISQWY 1269
              ::  |  :  ||: |: | :    : || |     ||          |   :    | ||
Db       1240 MEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDILIASNWY 1290

Qy       1270 LRRISENINKLRLGCNWQFIPVDEGWT 1296
                : : |    |||:| |:| |||||
Db       1291 FNHLKDKI----LGCDWYFVPTDEGWT 1313


RESULT 5
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match           34.2%;  Score 2328.5;  DB 2;  Length 1268;
  Best Local Similarity  38.6%;  Pred. No. 4.2e-97;
  Matches  519;  Conservative  252;  Mismatches  450;  Indels  123;  Gaps   34;

Qy          1 MPVNIKNFNYNDPINNDDIIMME-PFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQF 59
              ||||| |||||||||| |: |: |: :     ||||| |:| :||:|||    | :|  |
Db          1 MPVNINNFNYNDPINNTTILYMKMPYYEDS-NKYYKAFEIMDNVWIIPERNIIGKKPSDF 59

Qy         60 NASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
                 : |      |||| || ||||||:||||:||||||||| |:|| ||: | :  |||
Db         60 YPPISLDSGS-SAYYDPNYLTTDAEKDRFLKTVIKLFNRINSNPAGQVLLEEIKNGKPYL 118

Qy        120 GNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD--NFTDSMIM 177
              ||  |  ::| ||  : |:  |    |  |    :: ||:| |||| : :   |   :
Db        119 GNDHTAVNEFCANNRSTSVEIK-ESNGTTD---SMLLNLVILGPGPNILECSTFPVRIFP 174

Qy        178 N--GHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHV 235
```

```
Db        175 NNIAYDPSEKGFGSIQLMSFSTEYEYAFN---DNTD-------LFIADPAISLAHELIHV 224

Qy        236 LHGLYGIK-ISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQ 294

Db        225 LHGLYGAKGVTNKKVIEVDQGALMAAEKDIKIEEFITFGGQDLNIITNSTNQKIYVILLS 284

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354

Db        285 NYTAIASRLSQV-NRNNSALNTTYYKNFFQWKYGLDQDSNGNYTVNISKFNAIYKKL-FS 342

Qy        355 FTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKA 414

Db        343 FTECDLAQKFQVKNRSNYLFHF-KPFRLLDLLDDNIYSISEGFNIGS--LRVNNNGQNIN 399

Qy        415 VNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKA 474

Db        400 LNSRIVGPIPDNGLVERFVGLCKSIVSKKGTKNSLCIKVNNRDLFFVASESSYNENGINS 459

Qy        475 ETIAYNTQNNTIENN---FSIDQLILDNDLSSGIDLPNENTEPF--TNFDDIDIPVYIKQ 529

Db        460 PK---EIDDTTITNNNYKKNLDEVILDYNSDA---IPNLSSRLLNTTAQNDSYVPKYDSN 513

Qy        530 --SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKAN 587

Db        514 GTSEIKEYTVDKLNVFFYLYAQKAPEGESAISLTSSVNTALLDASKVYTFFSSDFINTVN 573

Qy        588 TVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNA 647

Db        574 KPVQAALFISWIQQVINDFTTEATQKSTIDKIADISLIVPYVGLALNIGNEVQKGNFKEA 633

Qy        648 FEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGL 704

Db        634 IELLGAGILLEFVPELLIPTILVFTIKSFINSDDSKNKIIKAINNALRERELKWKEVYSW 693

Qy        705 IVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFN--DID 762

Db        694 IVSNWLTRINTQFNKRKEQMYQALQNQVDGIKKIIEYKYNNYTLDEKNRLRAEYNIYSIK 753

Qy        763 FKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYL 822

Db        754 EELNKKVSLAMQNIDRFLTESSISYLMKLINEAKINKLSEYDKRVNQYLLNYILENSSTL 813

Qy        823 -LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLI 881

Db        814 GTSSVPELNNLVSNTLNNSIPFELSEYTNDKILIHILIRFYKRIIDSSILNMKYENNRFI 873

Qy        882 DSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRT 941

Db        874 DSSGYGSNISINGDIYIYSTNRNQFGIYSSRLSEVNITQNNTIIYNSRYQNFSVSFWVRI 933

Qy        942 PKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKD 1000

Db        934 PKYNN---LKNLNNEYTIINCMRNNNSGWKISLNYNNIIWTLQDTTGNNQKLVFNYTQMI 990

Qy       1001 NISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTK 1060

Db        991 DISDYINKWTFVTITNNRLGHSKLYINGNLTDQKSILNLGNIHVDDNILFKIVGCND-TR 1049

Qy       1061 FVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK- 1119

Db       1050 YVGIRYFKIFNMELDKTEIETLYHSEPDSTILKDFWGNYLLYNKKYYLLNLLKPNMSVTK 1109

Qy       1120 ------------YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNI 1167

Db       1110 NSDILNINRQRGIYSKTNI--------FSNA-----RLYTGVEVIIRKVGSTDTSNTDNF 1156

Qy       1168 VREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTY 1227

Db       1157 VRKNDTVYINVVDGNSE-YQLYADVSTSAVEKTIKLRRISN----------------SNY 1199

Qy       1228 NCQIL---------CEKDTKT-----FGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRI 1273

Db       1200 NSNQMIIMDSIGDNCTMNFKTNNGNDIGLLGF------------HLNNLVASSWYYKNI 1246

Qy       1274 SENINKLRLGCNWQFIPVDEGWTE 1297

Db       1247 RNNTR--NNGCFWSFISKEHGWQE 1268
```

RESULT 6
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.,
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin,
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arour
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>

R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 33.6%; Score 2291; DB 1; Length 1296;
Best Local Similarity 39.7%; Pred. No. 2.1e-95;
Matches 532; Conservative 216; Mismatches 502; Indels 90; Gaps 29;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||    | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: :
Db          1 MPFVNKQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGD 58

Qy         61 ASTGVFSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
               :    :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||:
Db         59 LNPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFW 118

Qy        120 GNAS--TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM 177
              | ::  |        |  ||      |||    : :  |  ||:| ||    :|
Db        119 GGSTIDTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKS 167

Qy        178 NGHSPIS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
               ||  ::    |:|:  ||| |      |     |     :    : |||:|| |||||
Db        168 FGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIH 227

Qy        235 VLHGLYGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIY 289
               | |||| |: |     ||  ::       |  ||| ||||||    |    :    : |
Db        228 AGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYY 287

Qy        290 NKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYK 349
              ||    |:|||: ||   |  |:   :   | ::| ||   || :||:|||| |||||||
Db        288 NK----FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYK 343

Qy        350 ALMFGFTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTE 407
               |   :|| |    : :  | :|  | : :  |     :: |||   :|||: : ||
Db        344 MLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAAN 400

Qy        408 FNGQNKAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNE 456
              |||||  :|   : ::          |::   : ::             |   :: || |||
Db        401 FNGQNTEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNW 460

Qy        457 DLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFT 516
              ||||  ::|:|: || | | |  :|     | | |:| ||    |:   |   ||
Db        461 DLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIEN 519

Qy        517 NFDDIDIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLND 566
               ||     | |  |           ||  :|  ::| || || |       : ||||:|:
Db        520 LSSDI-----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNE 574

Qy        567 ALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIII 626
              || | ::|||||::  |:|  |    |::|: ||: :: ||| |::: || ||::|::|||
Db        575 ALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIII 634

Qy        627 PYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMT 686
              ||||||||:||   |::|  |    || ||:|||||: :|::| | | ||: ||  :|
Db        635 PYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQT 694

Qy        687 ISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRY 746
              | ||| ||:|| ::|  ||: ||: ||||   |:::|  || ||::| : ||  |||:|
Db        695 IDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQY 754

Qy        747 SEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDN 806
              :||:| ||| : :|:  |||:||| |: ||: |:||||:||||| |||  ||:|:||| :
Db        755 TEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDAS 814

Qy        807 LKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNIS 866
              ||  ||:||  |   |: :|: || |||  |   ||| || |  :  |:  |  || ||
Db        815 LKDALLKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNII 874

Qy        867 SNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVY 926
              : :||:| |     ||| | | : :|:|| | |: |   | :| | |:| |       :||
Db        875 NTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVY 934

Qy        927 DSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVN 986
              :||::||| :||:| ||| |:     | ||||||:|::|:||||||:    |||||
Db        935 NSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQ 991

Qy        987 AKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSN 1046
                : : |:||   |||||||:|  :||||:|| |: |||||  |   : | ||  |::||
Db        992 EIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASN 1051

Qy       1047 DIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQY 1106
              :| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||||||: |:||  |
Db       1052 NIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPY 1111

Qy       1107 YLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRN 1161
              |: |    | |:  :   :  |     || :    |  :|  :|| | :||||| ::
Db       1112 YMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG-- 1168

Qy       1162 INNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-K 1220
              | ||||  | :|:|:  : :: ||:    :   ::  |    | |      |: :| |
Db       1169 -NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSK 1226

Qy       1221 YYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENI 1277
                :  | |:: | : :    |  | :|           :|      | || |:|
Db       1227 NDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLVASNWYNRQIER-- 1273

Qy       1278 NKLRLGCNWQFIPVDEGWTE 1297
              :   |||:|:|||||:|| |
Db       1274 SSRTLGCSWEFIPVDDGWGE 1293
```

RESULT 7
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1

A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

Query Match 33.1%; Score 2257; DB 2; Length 1252;
Best Local Similarity 38.5%; Pred. No. 6.8e-94;
Matches 513; Conservative 253; Mismatches 449; Indels 118; Gaps 35;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||  | :||||||:|:  |: ::|    |   :||:| |:  |||:|||   |  |  |:
Db          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                | : |   |||| ||::| |||:||| : |:|||||:  ||  ||: :  | ||||
Db         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNG 179
              | :|| ::|  :: : |  :     |::|    |: |:|| |  | | : | ::  : |
Db        116 NDNTPDNQF--HIGDASAVEIKFSNGSQDI---LLPNVIIMGAEPDLFETNSSNISLRNN 170

Qy        180 HSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              : | : |||:  :: | |     ||:   |:           : ||||||||||||| ||||
Db        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221

Qy        240 YGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKA 292
              || |      ||  ||     | :|:          ||  |||| | ::|: :   :||
Db        222 YGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275

Qy        293 LQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALM 352
              | ::: ||::|   |  | |   :: || ::: ||   :| :| |||: :||: ::| |
Db        276 LADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL- 331

Qy        353 FGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              : ||| :|| :: :| | :|  :|     |   ||:::||  :||:||   |||  | |||
Db        332 YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388

Qy        413 KAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLA 472
              :|      |:  ||   | || ::      :   || :|| :|||:|:::|:: |
Db        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNI 448

Qy        473 KA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYI 527
                    :|:   | ||  ||:  :||:||: :  |   | :|    |  :|  || |
Db        449 NTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYD 502

Qy        528 KQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEK 585
                  | :::  |:  ::| || ||  |    |: ||:|:: ||    |:|||||:  :
Db        503 SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINN 562

Qy        586 ANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFK 645
                                                                        |  | |:|||:|:: |: |||:|: ||||:||::|:||::|||| |||:|||  | |||
Db        563 VNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFK 622

Qy        646 NAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMY 702
              :| |: || ||:|| |||::| :  ||::|::|   ||  :|  |:||||:||:|| ::|
Db        623 DALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVY 682

Qy        703 GLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFND 760
                |||  |:: :||||    ||:|| || ||   ||: ||| :|| |: |:|  :   |
Db        683 SFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ 742

Qy        761 IDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNEL 820
              |: :||| :::|:|||| |: : ||||||  :  : : ||:::|:|:|  || ||  :
Db        743 IENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGS 802

Qy        821 YLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRL 880
              | :    | | |   | :|||| || || | |||  || :   | |:::|:: |:  :
Db        803 ILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKY 862

Qy        881 IDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVR 940
              :|:||| : :|:  ||          || : | : |  :   |: :::||: : ||||:|||
Db        863 VDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVR 922

Qy        941 TPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIK 999
              | |:|  :   : ||||||:|:: |:||||||:  | ||||| |    :: : | |
Db        923 IPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNA 980

Qy       1000 DNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTT 1059
              : |||||||  :|||||||:: :||||:|   : ||||  |: |::| ||::||:  |
Db        981 NGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCS-YT 1039

Qy       1060 KFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK 1119
              ::: |: |||| :||: ||: :||  : :|| ||||||| | || :||| |     | :|
Db       1040 RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFI- 1098

Qy       1120 YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 1179
                    |   |  :  :    || |::  |:: :||    |||:||: | :|:|
Db       1099 ---DRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN-- 1151

Qy       1180 NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKT--------TYNCQI 1231
                      | ||   | ||       | : |     |  |||           :|  :
Db       1152 -----------FVASK---THLF-------PLYADTATTNK--EKTIKISSSGNRFNQVV 1188

Qy       1232 L-------CEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGC 1284
              :       |  : |       ||      |:  ||        | ||   : :: |    ||
Db       1189 VMNSVGNNCTMNFKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDHTNS--NGC 1238

Qy       1285 NWQFIPVDEGWTE 1297
                | ||  : || |
Db       1239 FWNFISEEHGWQE 1251
```

RESULT 8
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 32.8%; Score 2233.5; DB 2; Length 1251;
Best Local Similarity 38.7%; Pred. No. 7.7e-93;
Matches 511; Conservative 246; Mismatches 469; Indels 95; Gaps 33;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||  | :||||||:||  |: ::|    |   :||:| |:  |||:|||   |  | |
Db          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                | : : |   |||| ||::| ||||||| : |:|||||   ||: ||: :  | ||||
Db         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115

Qy        121 NASTPPDKFAANVAN-VSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMN 178
              | :||   |  | |: | |       |::     |: |:|| |  | : | ::  : |
Db        116 NDNTPDGDFIINDASAVPIQ---FSNGSQSI---LLPNVIIMGAEPDLFETNSSNISLRN 169

Qy        179 GHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHG 238
              : | : |||:  :: | |      |      || |:     :  ||||||||||| |||
Db        170 NYMPSNHGFGSIAIVTFSPEYSFRF------KDNSM---NEFIQDPALTLMHELIHSLHG 220

Qy        239 LYGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNK 291
              |||  |     ||  ||    |  :|:          ||  |||| | ::|: :   :||
Db        221 LYGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTN 274

Qy        292 ALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKAL 351
              | ::: ||::|   |  | |   :: || ::: ||   :| :| |||: :||: ::| |
Db        275 LLADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL 331

Qy        352 MFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQ 411
              : ||| :|| :: :| | :|  :|    |   ||:::||  :||:|   |||  | ||
Db        332 -YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQ 387

Qy        412 NKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDL 471
              |  :|     |:   ||   |  || ::     :   || :|| :|||:|::|:: |
Db        388 NANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDN 447

Qy        472 AKA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVY 526
                    :|:   | ||  ||:  :||:||: :  |  | :|     |  :|  || |
Db        448 INTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKY 501

Qy        527 IKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVE 584
                  | :::  |:  ::| || ||  |    |: ||:|:: ||    |:|||||:  :
Db        502 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 561

Qy        585 KANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENF 644
              |  | |:||| |:: |: ||:|:|: ||||:||::|:|:::|||| |||:|||  | ||
Db        562 NVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 621

Qy        645 KNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDM 701
              |:| |: || ||:|| |||::| :  ||::|::|   ||  :|  |:||||:||:|| ::
Db        622 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 681

Qy        702 YGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFN 759
              |  ||| |:: :|||   ||:|| || ||  |:: ||| :|| |: |:|  :    |
Db        682 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIE 741

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819
              |: :||| :::|:|||| |: : |||||  :  : : ||:::|:|:|  ||:||  :
Db        742 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHG 801

Qy        820 LYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGR 879
                |  :    | | |   | :|||| || || | |||  || :   | |:::|:: |:  :
Db        802 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 861

Qy        880 LIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWV 939
              :|:||| : :|:  ||         || : | : | :   |: :::||: : ||||:|||
Db        862 YVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWV 921

Qy        940 RTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSI 998
              | | |:|  :   : ||||||:|:: |:||||||:  | ||||| | :  :: : | |
Db        922 RIPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGN 979

Qy        999 KDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDT 1058
              : |||||||  :|||||||:: :||||:|   : ||||  |: |::| ||::||:
Db        980 ANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCS-Y 1038

Qy       1059 TKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYI 1118
              |::: |: |||| :||: ||: :||  : : | ||||||  | || :||| |    | :|
Db       1039 TRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFI 1098

Qy       1119 KYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI 1178
                          |   |  :  :   || |::  |:: :||    |||:||: | :|:|
Db       1099 ----NRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN- 1151

Qy       1179 DNISDESYRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD 1236
                          | ||      |:   |  | : |         :: :    |    | :
Db       1152 ------------FVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMN 1199

Qy       1237 TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 1296
                |      ||       |:  ||        | ||   : :| |    |  | ||  : ||
Db       1200 FKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQ 1249

Qy       1297 E 1297
                |
Db       1250 E 1250
```

RESULT 9
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 32.4%; Score 2211; DB 2; Length 1296;
Best Local Similarity 38.1%; Pred. No. 8.3e-92;
Matches 515; Conservative 214; Mismatches 511; Indels 110; Gaps 28;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||   | ||| ||:|  ||  ::   :  |     |||:|  ::||::||| |:   |:: :
Db          1 MPFVNKQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGD 58

Qy         61 ASTGVFSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
                :   :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||:
Db         59 LNPPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFW 118

Qy        120 GNAS--TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM 177
              | ::  |      |  ||      |||    : : |  ||:| ||    :|
Db        119 GGSTIDTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKS 167

Qy        178 NGHSPIS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
                ||  ::    |:|:  ||| |     |    |     :     : |||:|| |||||
```

```
Db        168 FGHDVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIH 227

Qy        235 VLHGLYGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIY 289

Db        228 AEHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYY 287

Qy        290 NKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYK 349

Db        288 NK----FKDVASTLNKAKSIIGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYK 343

Qy        350 ALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFN 409

Db        344 MLTEIYTEDNFVNFFKVINRKTYLN-FDKAVFRINIVPDENYTIKDGFNLKGANLSTNFN 402

Qy        410 GQNKAVNKEAYEEIS-----LEHLVIYRIAMCKPVMYK------NTGKSEQCIIVNNEDL 458

Db        403 GQNTEINSRNFTRLKNFTGLFEFYKLLCVRGIIPFKTKSLDEGYNKALNDLCIKVNNWDL 462

Qy        459 FFIANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQL-------ILDND--------LSS 503

Db        463 FFSPSEDNFTNDLDKVEEITADTNIEAAEENISLDLIQQYYLTFDFDNEPENISIENLSS 522

Qy        504 GID---LPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQL 560

Db        523 DIIGQLEPMPNIERFPNG-------------KKYELDKYTMFHYLRAQEFEHGDSRIIL 568

Qy        561 TNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVS 620

Db        569 TNSAEEALLKPNVAYTFFSSKYVKKINKAVEAFMFLNWAEELVYDFTDETNEVTTMDKIA 628

Qy        621 DVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNK 680

Db        629 DITIIVPYIGPALNIGNMLSKGEFVEAIIFTGVVAMLEFIPEYALPVFGTFAIVSYIANK 688

Qy        681 GHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIE 740

Db        689 VLTVQTINNALSKRNEKWDEVYKYTVTNWLAKVNTQIDLIREKMKKALENQAEATKAIIN 748

Qy        741 DQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKL 800

Db        749 YQYNQYTEEEKNNINFNIDDLSSKLNESINSAMININKFLDQCSVSYLMNSMIPYAVKRL 808

Qy        801 KDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNN 860

Db        809 KDFDASVRDVLLKYIYDNRGTLVLQVDRLKDEVNNTLSADIPFQLSKYVDNKKLLSTFTE 868

Qy        861 YISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQ 920

Db        869 YIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLESSTIEVIL 928

Qy        921 SKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIW 980

Db        929 KNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSLNYGEIIW 985

Qy        981 TLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLD 1040

Db        986 TLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQKPISNLG 1045

Qy       1041 RINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPL 1100

Db       1046 NIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILKDFWGNYL 1105

Qy       1101 RYDTQYYLFNQGMQNIYIK---------YFSKASMGETAPRTNFNNAAINYQNLYLGLRF 1151

Db       1106 QYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNS-----TLYEGTKF 1160

Qy       1152 IIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPT 1211

Db       1161 IIKKYASG---NEDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGN 1216

Qy       1212 FYDVLQIKKYYEKTTYN-CQILCEKDTKTFGLFGIGKFVKDYGYV-WDTYDNY--FCISQ 1267

Db       1217 LSQVVVMKSKDDQGIRNKCKMNLQDNNG-----------NDIGFIGFHLYDNIAKLVASN 1265
```

```
Qy       1268 WYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1266 WYNRQVGK--ASRTFGCSWEFIPVDDGWGE 1293
```

RESULT 10
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 32.1%; Score 2188.5; DB 2; Length 1274;
Best Local Similarity 37.2%; Pred. No. 8.3e-91;
Matches 498; Conservative 258; Mismatches 477; Indels 105; Gaps 35;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMME-PFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQF 59

Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSK-KYYKAFEIMRNVWIIPERNTIGTNPSDF 59

Qy         60 NASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119

Db         60 DPPASL-KNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYL 118

Qy        120 GNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDN-------FT 172

Db        119 GNDHTPIDEFSPVTRTTSVNIKL----STNVESSMLLNLLVLGAGPDIFESCCYPVRKLI 174

Qy        173 DSMIMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHEL 232

Db        175 DPDVV--YDPSNYGFGSINIVTFSPEYEYTFNDISGGHNS---STESFIADPAISLAHEL 229

Qy        233 IHVLHGLYG---------IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPS 283

Db        230 IHALHGLYGARGVTYEETIEVKQAPL--------MIAEKPIRLEEFLTFGGQDLNIITSA 281

Qy        284 TDMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDK 343

Db        282 MKEKIYNNLLANYEKIATRLSEVNSAPPE-YDINEYKDYFQWKYGLDKNADGSYTVNENK 340

Qy        344 FDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKN 403

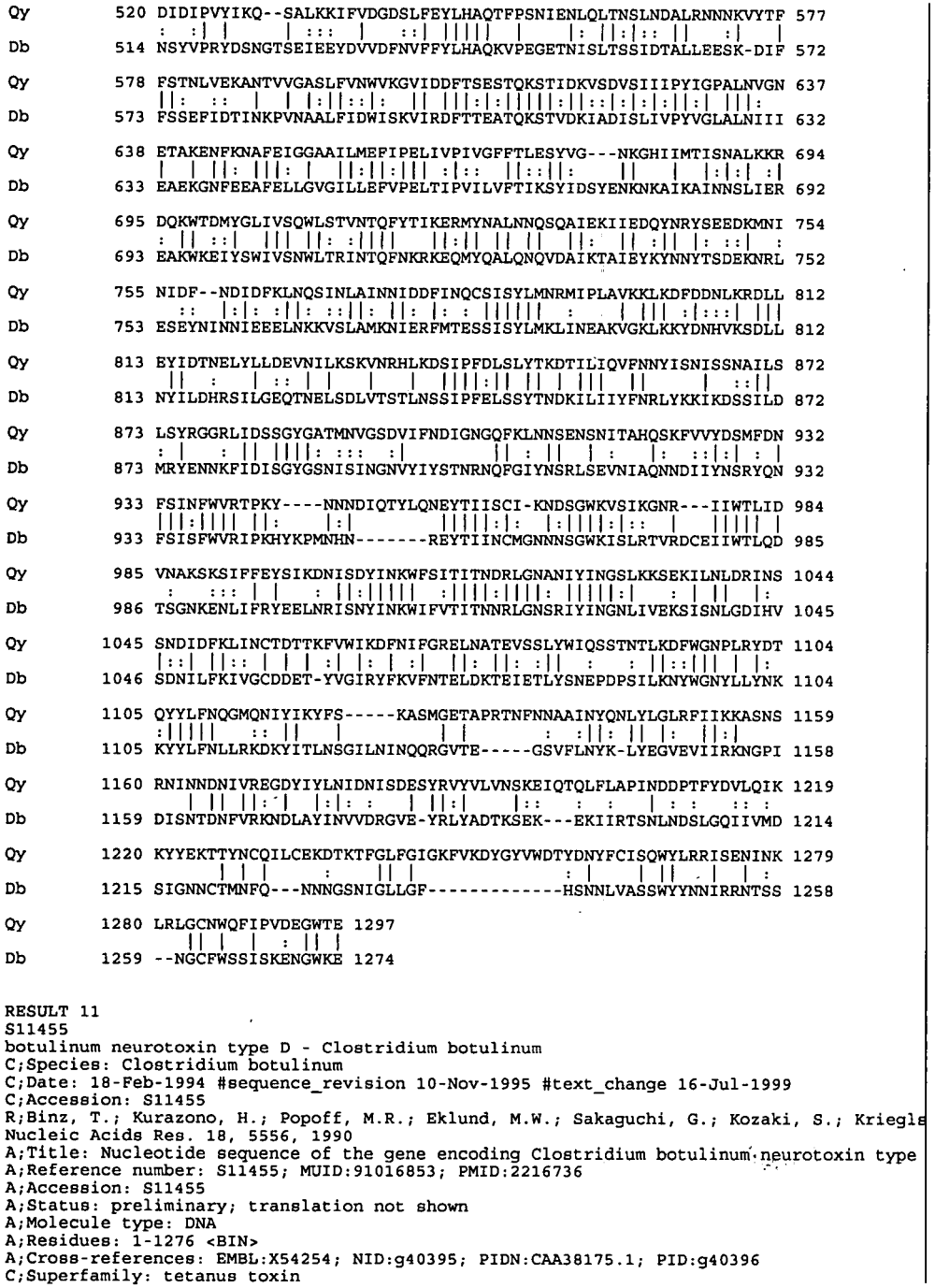
Db        341 FNEIYKKL-YSFTESDLANKFKVKCRNTYFIKY-EFLKVPNLLDDDIYTVSEGFNIG--N 396

Qy        404 LKTEFNGQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKS--EQCIIVNNEDLFFI 461

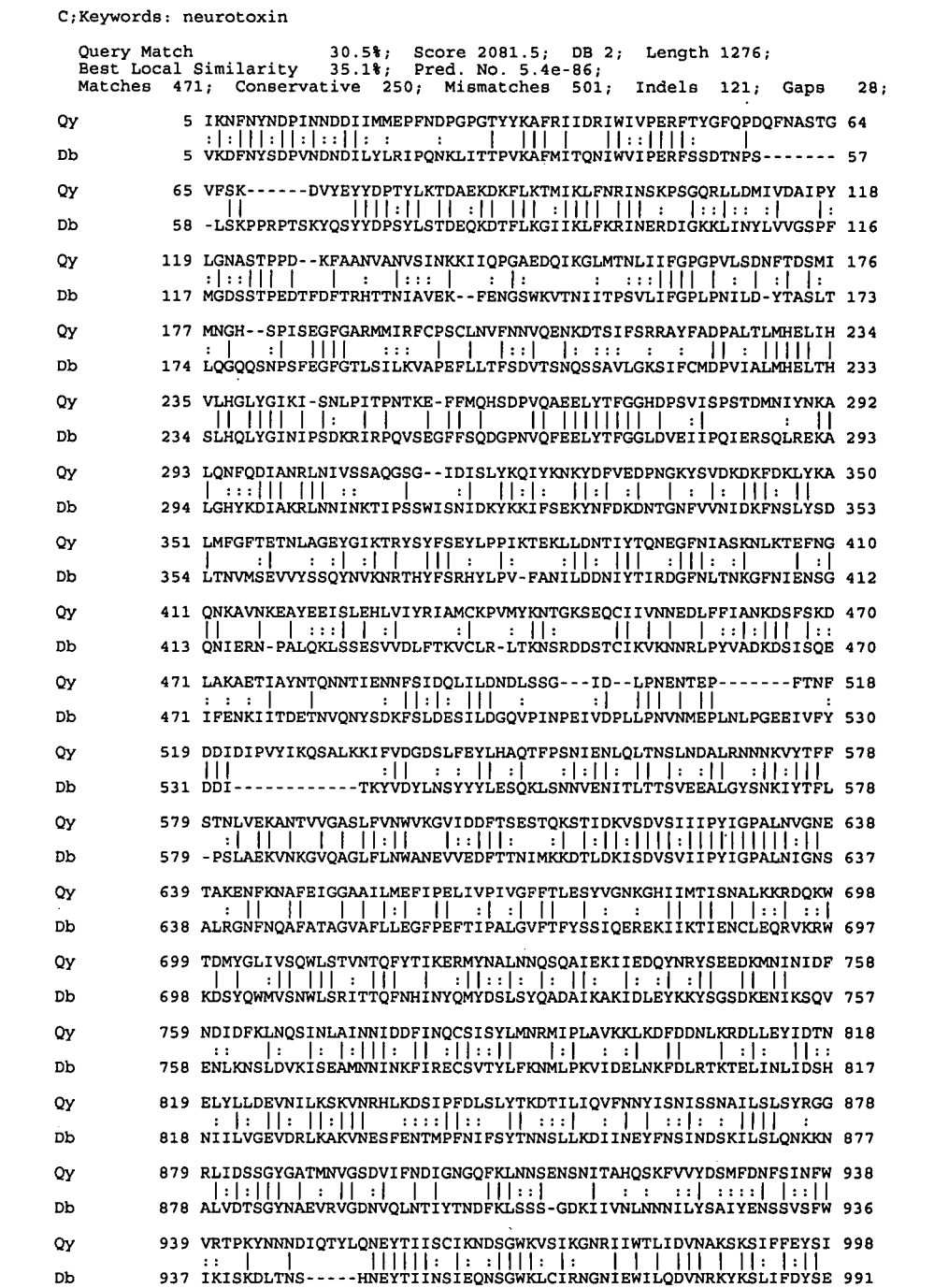
Db        397 LAVNNRGQSIKLNPKIIDSIPDKGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNSELFFV 456

Qy        462 ANKDSFSK-DLAKAETIAYNTQ-NNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFD 519

Db        457 ASESSYNENDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSQTIPQISNRTLNTLVQ-D 513
```



RESULT 11
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin



C;Keywords: neurotoxin

Query Match 30.5%; Score 2081.5; DB 2; Length 1276;
Best Local Similarity 35.1%; Pred. No. 5.4e-86;
Matches 471; Conservative 250; Mismatches 501; Indels 121; Gaps 28;

```
Qy        999 KDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDT 1058
Db        992 SLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDE 1051

Qy       1059 TKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYI 1118
Db       1052 NQMLWIRDFNIFSKELSNEDINIVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYI 1111

Qy       1119 KYFSKASMGETAPRTNFNNAAINY---QNLYLGLRFIIKKASN----SRNINNDNIVREG 1171
Db       1112 -----------APESNV-LVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNII--- 1156

Qy       1172 DYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQI 1231
Db       1157 ---------------LHMLYNSRKYMI------IRDTDTIY-ATQGGECSQNCVYALKL 1193

Qy       1232 LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--------------FCISQWYLRRISEN 1276
Db       1194 QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTN 1253

Qy       1277 INKLRLGCN--WQFIPVDEGWTE 1297
Db       1254 YETKLLSTSSFWKFISRDPGWVE 1276
```

RESULT 12
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membr
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match   29.8%; Score 2028; DB 2; Length 1285;
  Best Local Similarity 35.4%; Pred. No. 1.4e-83;
  Matches 475; Conservative 251; Mismatches 504; Indels 110; Gaps   33;

```
Qy          5 IKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 64
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPS------- 57

Qy         65 VFSK------DVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPY 118
Db         58 -LSKPPRPTSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPF 116

Qy        119 LGNASTPPD--KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMI 176
Db        117 MGDSSTPEDTFDFTRHTTNIAVEK--FENGSWKVTNIITPSVLIFGPLPNILD-YTASLT 173

Qy        177 MNGH--SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
Db        174 LQGQQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTH 233

Qy        235 VLHGLYGIKI-SNLPITPNTKE-FFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKA 292
Db        234 SLHQLYGINIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGSDVEIIPQIERLQLREKA 293

Qy        293 LQNFQDIANRLN----IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLY 348
Db        294 LGHYKDIAKRLNNINKTIPSSWSSNID--KYKKIFSEKYNFDKDNTGNFLVNIDKFNSLY 351

Qy        349 KALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEF 408
Db        352 SDLTNVMSEVIYSSQYNVKNRTHYFSKHYLPV-FANILDDNIYTIINGFNLTTKGFNIEN 410

Qy        409 NGQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFS 468
Db        411 SGQNIERN-PALQKLSSESVVDLFTKVCLR-LTRNSRDDSTCIQVKNNTLPYVADKDSIS 468

Qy        469 KDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSS---GID--LPNENTEPFTNFDDIDI 523
Db        469 QEIFESQIITDETNVENYSDNFSLDESILDAKVPTNPEAVDPLLPNVNMEP------LNV 522

Qy        524 PVYIKQSALKKIF-------VDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYT 576
Db        523 P------GEEEVFYDDITKDVDYLNSYYYLEAQKLSNNVENITLTTSVEEALGYSNKIYT 576

Qy        577 FFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVG 636
Db        577 FL-PSLAEKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSAIIPYIGPALNIG 635

Qy        637 NETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQ 696
Db        636 NSALRGNFKQAFATAGVAFLLEGFPEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVK 695

Qy        697 KWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINI 756
Db        696 RWKDSYQWMVSNWLSRITTQFNHISYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKS 755

Qy        757 DFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYID 816
Db        756 QVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLKTKTELINLID 815

Qy        817 TNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYR 876
Db        816 SHNIILVGEVDRLKAKVNESFENTIPFNIFSYTNNSLLKDMINEYFNSINDSKILSLQNK 875

Qy        877 GGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNS--ENSNITAHQSKFVVYDSMFDNFS 934
Db        876 KNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGSSGDDRGKIIVTQNENIVYNAMYESFS 935

Qy        935 INFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFF 994
Db        936 ISFWIRINKWVSN------LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNENSEQDINF 989

Qy        995 EYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK--- 1051
Db        990 SYDISKNAAGY-NKWFFVTITTNMMGNMMIYINGKLIDTIKVKELTGINFSKTITFQMNK 1048

Qy       1052 -----LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQY 1106
Db       1049 IPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSLQYTNVVKDYWGNDLRYDKEY 1108

Qy       1107 YLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
Db       1109 YMINVNYMN---RYMSKKGNG-IVFNTRKNNNDFNE-----GYKIIIKRI---RGNTNDT 1156

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTT 1226
Db       1157 RVRGENVLYFN-TTIDNKQYSLGMYKPSRNLGTDLVPLGALDQP----MDEIRKYGSFII 1211

Qy       1227 YNCQ----------ILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRIS-E 1275
Db       1212 QPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLG------DDYWFNHEYLIPVIKIE 1265

Qy       1276 NINKL--RLGCNWQFIPVDE 1293
Db       1266 HYASLLESTSTHWVFVPASE 1285
```

```
RESULT 13
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic p
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

  Query Match             28.0%;  Score 1908;  DB 2;  Length 1291;
  Best Local Similarity   34.2%;  Pred. No. 3.4e-78;
  Matches  464;  Conservative  240;  Mismatches  523;  Indels  130;  Gaps   32;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPF-----NDPGPGTYYKAFRIIDRIWIVPERFTYGFQ 55
              ||: | ||||:||::| :|: ::       |:|      |||||   ||::|:||:
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEP-----EKAFRITGNIWVIPDRFSRNSN 55

Qy         56 PDQFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDA 115
              |:  |    | |      |||| || ||::|| ||| :|||| ||||:  |: |:  :
Db         56 PN-LNKPPRVTSPK-SGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTD 113

Qy        116 IPYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSM 175
              ||:|| :|| : |  :|   |:: |  |      :  :  ::|| ||   : |  | :
Db        114 IPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF 173

Qy        176 -IMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
               : |      |||||  :|   |  :  ::|   :     ||:  :  || | |||||
Db        174 KLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNH 233

Qy        235 VLHGLYGIKISN----LPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYN 290
              :| |||| | |      :| |   |: |::  ::   |:| |||    :|  |
Db        234 AMHNLYGIAIPNDQTISSVTSNI--FYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEE 291

Qy        291 KALQNFQDIANRLNIVSSAQGSGID--ISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLY 348
              |||  :: || ||| :::|  |  :  |  |||    || || : :|: :|:::|| :||
Db        292 KALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELY 351

Qy        349 KALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEF 408
              |   ||| | |  | :: |  | |     |: |  :||: :|    ||||   ||   |
Db        352 NELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLF 410

Qy        409 NGQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNT--GKSEQC--IIVNNEDLFFIANK 464
              ||| : |  |  ::: |:::      |   :   :   |:  |  ::| | || || :
Db        411 MGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDI 469

Qy        465 DSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSS---------GID-----LPNE 510
              |:   : |   |:     :| |:||:||  : |            ||      | |
Db        470 SDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGE 529

Qy        511 NTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRN 570
              |   : |                :   ||  : : || :|    |:|:   | |: :|| |
Db        530 NQVFYDN--------------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDN 574

Qy        571 NNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIG 630
              : ||||:| | |  | |  |   ||: |   |::|||:   :| |:||:|||| ||||||
Db        575 SAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIG 633

Qy        631 PALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNA 690
              ||||: |   : ||  || : |  ||:|  ||  :| :| | : | |  :  || || |
Db        634 PALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNC 693

Qy        691 LKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEED 750
              |::| ::| | |  ::  ||| : |||  |  :||::|| |: ||:  |: :| :||  |
Db        694 LEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSD 753

Qy        751 KMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRD 810
              | ||     ::   |:  |: |:|||: || :||::||   |:|  : :| :|| | |
Db        754 KENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAK 813

Qy        811 LLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAI 870
              |:  ||:: : |: ||: ||:|||   :::|||::  || :::|  : | | :||: : |
Db        814 LINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKI 873

Qy        871 LSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNS--ENSNITAHQSKFVVYDS 928
              ||| |   |:|:||| | ::   ||  | |    ||| :|  :   :   |:: :||:|
Db        874 LSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNS 933

Qy        929 MFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAK 988
              |:::|||:||:|  |: :|         ||||  :||:||| : |  | :::||
Db        934 MYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDS 987

Qy        989 SKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDI 1048
              :|| | | | :|   | |||| :|:||: :||  |||||  |  : |:  |  || |  |
Db        988 EQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTI 1046

Qy       1049 DFKLINCTDTTKF--------VWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPL 1100
               |::   ||           :||:|| || :||:  ::: |:     || :||:||| |
Db       1047 TFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDL 1106

Qy       1101 RYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAAINYQNLYLGLRFIIKKASNS 1159
              ||: :||: |    | |:   |:  :  |    | ||           | : |||:
Db       1107 RYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE----------GYKIIIKRI--- 1153

Qy       1160 RNINNDNIVREGDYIYLNI-----------------DNISDES-YRVYVLVNSKEIQTQ 1200
              |   ||  || || :| ::                || | |  | : :   :|:|
Db       1154 RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDN 1213

Qy       1201 LF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDT 1258
              :  : |:|:  |:|   ||     |: :| : :         |:  || :    | |
Db       1214 IIFQIQPMNN--TYYYASQI----FKSNFNGENIS-------GICSIGTYRFRLGGDWYR 1260

Qy       1259 YDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 1293
              : ||    : |     : |: :       :| |:|| |
Db       1261 H-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1291


RESULT 14
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.,
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
```

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            28.0%;  Score 1908;  DB 2;  Length 1291;
  Best Local Similarity  34.2%;  Pred. No. 3.4e-78;
  Matches  464;  Conservative  240;  Mismatches  523;  Indels  130;  Gaps   32;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPF-----NDPGPGTYYKAFRIIDRIWIVPERFTYGFQ 55
              ||: | ||||:||::| :|: ::       |:|      |||||   ||::|:||:
Db          1 MPITINNFNYSDPVDNKNILYLDTHLNTLANEP-----EKAFRITGNIWVIPDRFSRNSN 55

Qy         56 PDQFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDA 115
              |:  |    | |      |||| || ||::|| ||| :|||| ||||:  |: |:  :
Db         56 PN-LNKPPRVTSPK-SGYYDPNYLSTDSDKDTFLKEIIKLKFKRINSREIGEELIYRLSTD 113

Qy        116 IPYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSM 175
              ||: || :|| : |  :|   |:: | |      :  :  ::|| ||   : |  | :
Db        114 IPFPGNNNTPINTFDPDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTF 173

Qy        176 -IMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
               : |       |||||  :|   |  :  ::|   :      ||:  :  || | |||| |
Db        174 KLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNH 233

Qy        235 VLHGLYGIKISN----LPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYN 290
              :| |||| | |       :| |   |: |::  ::  |:| |||    :|  |
Db        234 AMHNLYGIAIPNDQTISSVTSNI--FYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEE 291

Qy        291 KALQNFQDIANRLNIVSSAQGSGID--ISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLY 348
              |||  :: || ||| :::|  |  :  |  |||     || || : :|: :|:::|| :||
Db        292 KALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELY 351

Qy        349 KALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEF 408
              |   ||| | |  | :: |  | |     |: |  :||: :|    ||||   ||   |
Db        352 NELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLF 410

Qy        409 NGQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNT--GKSEQC--IIVNNEDLFFIANK 464
              ||| : |  |  ::: |:::      |  :  :   |:  |  ::| | || || :
Db        411 MGQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDI 469

Qy        465 DSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSS---------GID-----LPNE 510
              |:  : |   |:     :| |:||:||  : |             ||      || |
Db        470 SDVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGE 529

Qy        511 NTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRN 570
              |  : |                :  ||  : : || :|    |:|:   | |: :|| |
Db        530 NQVFYDN--------------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDN 574

Qy        571 NNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIG 630
              : ||||:| | |  | | |    ||: |   |::|||:   :| |:||:|||| ||||||
Db        575 SAKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIG 633

Qy        631 PALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNA 690
              ||||: |   : ||  || : |  ||:|  ||  :| :| | : | |  :  || || |
Db        634 PALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNC 693

Qy        691 LKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEED 750
              |::| ::| | |  ::  ||| : |||  |  :||::|| |: ||:  |: :| :||  |
Db        694 LEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSD 753

Qy        751 KMNININIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRD 810
```

```
              | ||      ::   |:  |: |:|||: || :||::||   |:|  : :| :|| | |
Db        754 KENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAK 813

Qy        811 LLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAI 870
              |:  ||:: : |: ||: ||:|||   :::|||::  || :::|  : | | :||: : |
Db        814 LINLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKI 873

Qy        871 LSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNS--ENSNITAHQSKFVVYDS 928
              |||  |   |:|:||| | ::   ||  | |    ||| :|  :   :   |:: :||:|
Db        874 LSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNS 933

Qy        929 MFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAK 988
              |:::|||:||:|  |: :|         ||||  :||:||| : |  | :::||
Db        934 MYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDS 987

Qy        989 SKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDI 1048
              :|| | | | :|   | |||| :|:||: :||  |||||  |  : |:  |  || |  |
Db        988 EQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTI 1046

Qy       1049 DFKLINCTDTTKF--------VWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPL 1100
              |::    ||           :||:|| || :||:  ::: |:      || :||:||| |
Db       1047 TFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDL 1106

Qy       1101 RYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAAINYQNLYLGLRFIIKKASNS 1159
              ||: :||: |    | |:   |:  :  |   | ||          | : |||:
Db       1107 RYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE----------GYKIIIKRI--- 1153

Qy       1160 RNINNDNIVREGDYIYLNI-----------------DNISDES-YRVYVLVNSKEIQTQ 1200
              |   ||  || || :| ::                 || | |  | : :   :|:|
Db       1154 RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDN 1213

Qy       1201 LF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDT 1258
              :   : |:|:  |:|   ||      |: :| : :         |:  || :   | |
Db       1214 IIFQIQPMNN--TYYYASQI----FKSNFNGENIS-------GICSIGTYRFRLGGDWYR 1260

Qy       1259 YDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 1293
              : ||   : |    : |: :      :| |:|| |
Db       1261 H-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1291


RESULT 15
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match           11.6%;  Score 791.5;  DB 2;  Length 367;
  Best Local Similarity  40.9%;  Pred. No. 9.3e-29;
  Matches  159;  Conservative   78;  Mismatches  121;  Indels   31;  Gaps    5;

Qy        641 KENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQK 697
              | |||:| |: || ||:|| |||::| :  ||::|::|   ||  :|  |:||||:||:|
Db          3 KGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEK 62

Qy        698 WTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--N 755
              | ::|  ||| |:: :||||   ||:|| || ||  ||: ||| :|| |: |:|  :
Db         63 WKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNK 122

Qy        756 IDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYI 815
                |  |: :||| :::|:|||  |: : |||||  :  : : ||:::|:|:|  || ||
```

```
Db        123 YDIKQIENELNQKVSIAMNNIYRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYI 182

Qy        816 DTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSY 875
              :   | :    | | |   | :|||| || || | ||| || :   | |:::|:: |
Db        183 IQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRY 242

Qy        876 RGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSI 935
              :  : :|:||| : :|:  ||         || : | : | :   |: :::||: : ||||
Db        243 KNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSI 302

Qy        936 NFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFF 994
              :|||| | |:|  :   : ||||||:|:: |:||||||:  | ||||| |    :: : |
Db        303 SFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAF 360

Qy        995 EYSIKDNISDYINKWFSITITNDRLGNAN 1023
              |                          ||||
Db        361 NY----------------------GNAN 366
```

Search completed: September 10, 2004, 03:44:21
Job time : 26.9134 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 15.8251 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 6816
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6804	99.8	1296	1	BXG_CLOBO	Q60393 clostridium
2	3992	58.6	1290	1	BXB_CLOBO	P10844 clostridium
3	2535.5	37.2	1314	1	TETX_CLOTE	P04958 clostridium
4	2285	33.5	1295	1	BXA1_CLOBO	P10845 clostridium
5	2237	32.8	1250	1	BXE_CLOBO	Q00496 clostridium
6	2232	32.7	1250	1	BXE_CLOBU	P30995 clostridium
7	2206	32.4	1295	1	BXA2_CLOBO	Q45894 clostridium
8	2188.5	32.1	1274	1	BXF_CLOBO	P30996 clostridium
9	2081.5	30.5	1276	1	BXD_CLOBO	P19321 clostridium
10	1903	27.9	1290	1	BXC1_CLOBO	P18640 clostridium
11	680.5	10.0	1162	1	BXEN_CLOBU	Q06366 clostridium
12	668.5	9.8	1162	1	BXEN_CLOBO	P46082 clostridium
13	619	9.1	1196	1	BXCN_CLOBO	P46081 clostridium
14	273.5	4.0	2710	1	TOXA_CLODI	P16154 clostridium
15	260	3.8	2136	1	YCF2_MARPO	P09975 marchantia
16	240	3.5	2366	1	TOXB_CLODI	P18177 clostridium
17	236	3.5	1169	1	EX5B_BORBU	O51578 borrelia bu
18	235.5	3.5	2339	1	RPC1_PLAFA	P27625 plasmodium
19	235	3.4	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
20	227.5	3.3	1225	1	Y309_MYCGE	P47551 mycoplasma
21	227.5	3.3	2867	1	RBP2_PLAVB	Q00799 plasmodium
22	226	3.3	1179	1	EX5B_BUCAP	Q8k9a9 buchnera ap
23	226	3.3	1197	1	DPOM_PODAN	Q01529 podospora a
24	216	3.2	2894	1	YD96_METJA	Q58791 methanococc
25	212	3.1	970	1	Y087_BUCAI	P57189 buchnera ap
26	211	3.1	1726	1	MSP1_PLAFC	P04934 plasmodium
27	211	3.1	1726	1	MSP1_PLAFP	P50495 plasmodium
28	210	3.1	3135	1	S230_PLAFO	Q08372 plasmodium
29	208.5	3.1	1386	1	RPOD_MARPO	P06274 marchantia
30	207.5	3.0	1956	1	ATX1_PLAFA	Q04956 plasmodium
31	203.5	3.0	1562	1	YM81_YEAST	Q04781 saccharomyc
32	203	3.0	1177	1	Y307_MYCGE	P47549 mycoplasma
33	201	2.9	2971	1	YCX9_CHLRE	Q32065 chlamydomon
34	198.5	2.9	1070	1	EX5C_BUCAI	P57528 buchnera ap
35	197.5	2.9	950	1	Y511_RICPR	Q9zd36 rickettsia
36	197.5	2.9	1024	1	RPOB_PLAFA	P21421 plasmodium
37	197.5	2.9	1174	1	EX5B_BUCAI	P57529 buchnera ap
38	196.5	2.9	968	1	Y080_BUCAP	Q8ka32 buchnera ap
39	196.5	2.9	1163	1	YPT4_CAEEL	P41882 caenorhabdi
40	196	2.9	997	1	T257_ECOLI	P25239 escherichia
41	196	2.9	1061	1	EX5C_BUCAP	Q8k9b0 buchnera ap
42	196	2.9	1630	1	MSP1_PLAFK	P04932 plasmodium
43	196	2.9	1639	1	MSP1_PLAFW	P04933 plasmodium
44	194.5	2.9	986	1	EPIB_STAEP	P30195 staphylococ
45	194	2.8	1111	1	EX5C_BUCBP	Q89ab4 buchnera ap

ALIGNMENTS

RESULT 1
BXG_CLOBO
ID BXG_CLOBO STANDARD; PRT; 1296 AA.
AC Q60393;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE (Bontoxilysin G).
GN BOTG.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;
RX MEDLINE=94092745; PubMed=8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins.";
RL Biochim. Biophys. Acta 1216:487-491(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; X74162; CAA52275.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.

```
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match           99.8%;  Score 6804;  DB 1;  Length 1296;
  Best Local Similarity   99.9%;  Pred. No. 3.6e-300;
  Matches 1295;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 PVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNA 61
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNA 60

Qy         62 STGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGN 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 STGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGN 120

Qy        122 ASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHS 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHS 180

Qy        182 PISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYG 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYG 240

Qy        242 IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIAN 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIAN 300

Qy        302 RLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLA 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNLA 360

Qy        362 GEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYE 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYE 420

Qy        422 EISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNT 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNT 480

Qy        482 QNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDS 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 QNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDS 540

Qy        542 LFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKG 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKG 600

Qy        602 VIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIP 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 VIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIP 660

Qy        662 ELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIK 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIK 720

Qy        722 ERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFIN 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFIN 780

Qy        782 QCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSI 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 QCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSI 840

Qy        842 PFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDI 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 PFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDI 900

Qy        902 GNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIIS 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIIS 960

Qy        962 CIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGN 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGN 1020

Qy       1022 ANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSS 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSS 1080

Qy       1082 LYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAIN 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 LYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAIN 1140

Qy       1142 YQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQL 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 YQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQL 1200

Qy       1202 FLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDN 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 FLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDN 1260

Qy       1262 YFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              ||||||||||||||||||||||||||||||||||||
Db       1261 YFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1296


RESULT 2
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
```

```
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             58.6%;  Score 3992;  DB 1;  Length 1290;
  Best Local Similarity   57.7%;  Pred. No. 3.8e-173;
  Matches  751;  Conservative  250;  Mismatches  283;  Indels   18;  Gaps   11;

Qy          2 PVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNA 61
              || | |||||||:|::||||||     | | |||||:| |||||:|||:|:|::|: ||
Db          1 PVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNK 60

Qy         62 STGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGN 121
               |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::||:||:: |||||:
Db         61 SSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLGD 120

Qy        122 ASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHS 181
                 | ::|  |:|:|::|| |  ||  :: ||:  ||||||||||::| |  : :  |
Db        121 RRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHF 180

Qy        182 PISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYG 241
                ||||  | ::||| ::|||||||||  |||:|| ||:|||| ||||||||||||||
Db        181 ASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLYG 240

Qy        242 IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIAN 301
              ||: :||| || |:|||| :| :|||||||||| |||:|:|||| :||:| ||||: | :
Db        241 IKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIVD 300

Qy        302 RLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL 360
              ||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:||||||||:
Db        301 RLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETNI 360

Qy        361 AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY 420
              |  | |||| |||:  |||:| : |||| |||  |||||: |::: |: |||||:||:||
Db        361 AENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQAY 420

Qy        421 EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN 480
              |||| ||| :|:| ||| |  |  |    || |:||||||||:|:||| ||:| | | ||
Db        421 EEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIEYN 475

Qy        481 TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD 540
              ||:| |||:| |::|||| || | |:||:|||  |:| ::|:||| || |:|||  | :
Db        476 TQSNYIENDFPINELILDTDLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDEN 534

Qy        541 SLPEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK 600
              ::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  |||
Db        535 TIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVK 594

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
               :::||  |: : :|:||::|:|:|:|||| |||||||||| ||:||||| ||:||:|||
Db        595 QIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEFI 654

Qy        661 PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI 720
              |||::|:|| | |||| || || || ||| ||::||:|||||||:||||||||||||||
Db        655 PELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTI 714
```

```
Qy        721 KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI 780
Db        715 KEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNFI 774

Qy        781 NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS 840
Db        775 NGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTI 834

Qy        841 IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND 900
Db        835 MPFDLSIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND 894

Qy        901 IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII 960
Db        895 --KNQFKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTII 952

Qy        961 SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG 1020
Db        953 NCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LN 1011

Qy       1021 NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS 1080
Db       1012 NAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIE 1071

Qy       1081 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA- 1139
Db       1072 ERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSK 1131

Qy       1140 -INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQ 1198
Db       1132 YINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEE 1189

Qy       1199 TQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYV 1255
Db       1190 EKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIV 1248

Qy       1256 WDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
Db       1249 FEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1290
```

RESULT 3
TETX_CLOTE
ID TETX_CLOTE STANDARD; PRT; 1314 AA.
AC P04958;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE chain (Tetanus toxin chain H)].
GN TETX OR CTP60.
OS Clostridium tetani.
OG Plasmid pE88, and Plasmid 75 Kbp.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC PLASMID=75 Kbp;
RX MEDLINE=87053814; PubMed=3536478;
RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA Weller U., Hudel M., Habermann E., Niemann H.;
RT "Tetanus toxin: primary structure, expression in E. coli, and
RT homology with botulinum toxins.";
RL EMBO J. 5:2495-2502(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CN3911; PLASMID=75 Kbp;
RX MEDLINE=87040747; PubMed=3774547;
RA Fairweather N.F., Lyness V.A.;
RT "The complete nucleotide sequence of tetanus toxin.";
RL Nucleic Acids Res. 14:7809-7812(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88; PLASMID=pE88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN [4]
RP SEQUENCE OF 742-1314 FROM N.A.
RC PLASMID=75 Kbp;
RX MEDLINE=86085672; PubMed=3510187;
RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT fragment C in Escherichia coli.";
RL J. Bacteriol. 165:21-27(1986).
RN [5]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=90201034; PubMed=2108021;
RA Krieglstein K., Henschen A., Weller U., Habermann E.;
RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT in tetanus toxin.";
RL Eur. J. Biochem. 188:39-45(1990).
RN [6]
RP PARTIAL SEQUENCE.
RX MEDLINE=92037649; PubMed=1935979;
RA Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT "Limited proteolysis of tetanus toxin. Relation to activity and
RT identification of cleavage sites.";
RL Eur. J. Biochem. 202:41-51(1991).
RN [7]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93010948; PubMed=1396558;
RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA Montecucco C.;
RT "Tetanus toxin is a zinc protein and its inhibition of
RT neurotransmitter release and protease activity depend on zinc.";
RL EMBO J. 11:3577-3583(1992).
RN [8]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RL Nature 359:832-835(1992).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE=97475217; PubMed=9334741;
RA Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA Sax M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin.";
RL Nat. Struct. Biol. 4:788-792(1997).
CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC   RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC   AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC   WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC   INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC   ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC   BOND OF SYNAPTOBREVIN-2.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC   synaptobrevin 2.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC   YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC   AND ARE NON-TOXIC AFTER SEPARATION.
CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC   GANGLIOSIDE RECEPTORS.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC --

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190
```

Query Match 37.2%; Score 2535.5; DB 1; Length 1314;
Best Local Similarity 40.0%; Pred. No. 2.4e-107;
Matches 539; Conservative 253; Mismatches 469; Indels 85; Gaps 25;

```
Qy          2 PVNIKNFNYNDPINNDDIIMMEPPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNA 61
              |: | || |:||:||| ||||||    |   |||||:| |||||||||: :| :|: ||
Db          1 PITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFNP 60

Qy         62 STGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGN 121
                : :  :   ||||| ||:||::||:||:||:|||||| :  :|: ||| |::||||||
Db         61 PSSLI-EGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN 119

Qy        122 ASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM---- 177
              : :  |||  |  :|| |      |         ::|||||||||||: |   :::
Db        120 SYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDN 179

Qy        178 NGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLH 237
              : |  :|||: | : |||  :  |:|| ||  :    :  || |||| ||||||||||
Db        180 KNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLH 239

Qy        238 GLYGIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQ 297
              ||||:::|:  | |: :| :|||: |: ||||:|||| | ::||     ::| | | :::
Db        240 GLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDYK 299

Qy        298 DIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTE 357
              |||:|: |:|     |||  |||||: || | :| ||:| |::|||  || ::|:||||
Db        300 AIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFTE 359

Qy        358 TNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNK 417
              |  :: |||| ||||    |:|   |||:|||   ||||| ||:||:|: |||  ||
Db        360 IELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVNT 419

Qy        418 EAYEEISLEHLVIYRIAMCKPV---------MYKNTGK-----SEQCIIVNNEDLFFIAN 463
              |:  :    ||   | :|| :          :|  |          | || : |||| |||
Db        420 NAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFIAE 479

Qy        464 KDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDI 523
              |:|||::  : | ::|||:|  :  |:|:|::|:| :| | | |||: | | |    |
Db        480 KNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTK--GIPY 537

Qy        524 PVYIKQSALKKI---FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFST 580
                | :|   |    :| ::::::||:||  |: :: : :|||::||| |: |:|::|
Db        538 APEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYF-P 596

Qy        581 NLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETA 640
              ::: | |       ||: ||: :|||||:||:||:||||:|||| |:|||||||:  :
Db        597 SVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQGY 656
```

```
Qy        641 KENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTD 700
              :  ||  | |   |  :|:|:|||: :|::   ::       |  || || | |:|| :|| :
Db        657 EGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWIE 716

Qy        701 MYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFND 760
              :| |: ::|| ||||||     :|| :|  |  ||:|||: :|  ||  ||  |  : |:
Db        717 VYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEINN 776

Qy        761 IDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNEL 820
              :  || :  | |: ||: |: : | |:|:|:||  | |:| :||    |   |::||  |
Db        777 LKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANSK 836

Qy        821 YL-LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN-------ISSNAILS 872
              :: : |:  |:||:|:     |||     |:|:        : :: |       :  : ||:
Db        837 FIGITELKKLESKINKVFSTPIPFS---YSKN------LDCWVDNEEDIDVILKKSTILN 887

Qy        873 LSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
              |       : | ||: :::    |  :    |       | |:|:|  :  |::  : |: ||:
Db        888 LDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMFN 947

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRIIWTLIDVN 986
              ||:::||:| || : : :: |  |||:||| :|       ||| ||:||| :|||| |
Db        948 NFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTLKDSA 1007

Qy        987 AKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSS 1045
               : : | |   :  |  : |: |||  |||||||| :||:|||| |  | :|  |  |
Db       1008 GEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIRED 1066

Qy       1046 NDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQ 1105
              |:|  ||  | :  ::| |  | || : ||  |:  ||    |   |:||||||||||:
Db       1067 NNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDTE 1126

Qy       1106 YYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFIIKKASNSRN 1161
              |||       .::: :|  :       ||  :: |  :|  |: || ||:|||: : :
Db       1127 YYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFIIKRYTPNNE 1184

Qy       1162 INNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPINDDPTFYDVL 1216
              |  |: |: ||:| | :   ::|    |       |   |.  : ||      |  :
Db       1185 I--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP---GIPLYKKM 1239

Qy       1217 QIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNYFCISQWYL 1270
              :  |  :  ||: |: | :    : || |     ||          |   :    | ||
Db       1240 EAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDILIASNWYF 1290

Qy       1271 RRISENINKLRLGCNWQFIPVDEGWT 1296
                : : |    |||:| |:| |||||
Db       1291 NHLKDKI----LGCDWYFVPTDEGWT 1312
```

RESULT 4
BXA1_CLOBO
ID BXA1_CLOBO STANDARD; PRT; 1295 AA.
AC P10845; P01561; P18639;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 2916;
RX MEDLINE=90235864; PubMed=2185020;
RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type A
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 189:73-81(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=62A;
RX MEDLINE=90264400; PubMed=2160960;
RA Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [3]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=62A;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN=Hall;
RX MEDLINE=89350959; PubMed=2669749;
RA Betley M.J., Somers E., Dasgupta B.R.;
RT "Characterization of botulinum type A neurotoxin gene: delineation of
RT the N-terminal encoding region.";
RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN [5]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=Type A NIH;
RX MEDLINE=96096783; PubMed=8521962;
RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT components of Clostridium botulinum type A progenitor toxins.";
RL FEBS Lett. 376:41-44(1995).
RN [6]
RP SEQUENCE OF 1-16.
RX MEDLINE=84178501; PubMed=6370252;
RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequence of the heavy and light chains of
RT botulinum neurotoxin type A.";
RL Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN [7]
RP SEQUENCE OF 1-46.
RA Dasgupta B.R., Foley J., Niece R.;
RT "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL Biochemistry 26:4162-4162(1987).
RN [8]
RP SEQUENCE OF 1-5 AND 444-456.
RX MEDLINE=91120847; PubMed=2126206;
RA Dasgupta B.R., Dekleva M.L.;
RT "Botulinum neurotoxin type A: sequence of amino acids at the
RT N-terminus and around the nicking site.";
RL Biochimie 72:661-664(1990).
RN [9]
RP SEQUENCE OF 448-464 AND 872-895.
RX MEDLINE=89024662; PubMed=3178218;
RA Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT halves and their partial sequences.";
RL Arch. Biochem. Biophys. 266:142-151(1988).
RN [10]
RP SEQUENCE OF 448-482.
RX MEDLINE=85285016; PubMed=3896784;
RA Shone C.C., Hambleton P., Melling J.;
RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT and purification of two tryptic fragments. Proteolytic action near
RT the COOH-terminus of the heavy subunit destroys toxin-binding
RT activity.";
RL Eur. J. Biochem. 151:75-82(1985).
RN [11]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;

```
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match           33.5%;  Score 2285;  DB 1;  Length 1295;
  Best Local Similarity  39.7%;  Pred. No. 4.9e-96;
  Matches  530;  Conservative  216;  Mismatches  499;  Indels   90;  Gaps   29;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65
              | ||| ||:|  ||  ::   : |     |||:| ::||::||| |:   |:: : :
Db          5 KQFNYKDPVNGVDIAYIK-IPNVGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 62

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123
               :| |   ||| ||| || ||| :|| : ||| || |   |: ||  ||  ||: | ::
Db         63 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 122

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182
               |       |  ||      |||    : : |  ||:| ||    :|          ||
Db        123 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHEV 171

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              ::    |:|:   ||| |      |    |      :      :  |||:|| |||||  | |
Db        172 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRL 231

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294
              ||| |: |     ||  ::      |  ||| ||||||   |    :    :  |||
Db        232 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 288

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
               |:|||: ||    |  |:   :   | ::| ||    || :||:|||| ||||||| |
Db        289 -FKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 347

Qy        355 FTETNLAGEYGIKTRYSY--FSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              :|| |    : :  | :|  | : :  |      :: |||   :|||: : ||   |||||
Db        348 YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIY---DGFNLRNTNLAANFNGQN 404

Qy        413 KAVNKEAYEEIS--LEHLVIYRIAMCKPVMYK---------NTGKSEQCIIVNNEDLFFI 461
              :|   : ::       |::   : ::          |   :: || ||| ||||
Db        405 TEINNMNFTKLKNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFS 464

Qy        462 ANKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDI 521
              ::|:|: || | | |  :|     | | |:| ||    |:   |   ||         ||
Db        465 PSEDNFTNDLNKGEEITSDTNIEAAEENISLD-LIQQYYLTFNFDNEPENISIENLSSDI 523
```

```
Qy        522 DIPVYIKQSAL----------KKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571
Db        524 -----IGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNP 578

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631
Db        579 SRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGP 638

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691
Db        639 ALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNAL 698

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751
Db        699 SKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEK 758

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811
Db        759 NNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDAL 818

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871
Db        819 LKYIYDNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSIL 878

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFD 931
Db        879 NLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYE 938

Qy        932 NFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKS 991
Db        939 NFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR 995

Qy        992 IFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK 1051
Db        996 VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFK 1055

Qy       1052 LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQ 1111
Db       1056 LDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNL 1115

Qy       1112 GMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
Db       1116 YDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDN 1171

Qy       1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKT 1225
Db       1172 IVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGI 1230

Qy       1226 TYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRL 1282
Db       1231 TNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTL 1277

Qy       1283 GCNWQFIPVDEGWTE 1297
Db       1278 GCSWEFIPVDDGWGE 1292
```

```
RESULT 5
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             32.8%;  Score 2237;  DB 1;  Length 1250;
  Best Local Similarity   38.3%;  Pred. No. 7e-94;
  Matches  508;  Conservative  253;  Mismatches  451;  Indels  116;  Gaps   34;

Qy          5 IKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 64
              | :||||||:|:  |: ::|     |   :||:| |:  |||:|||    | |  |: |
Db          3 INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS 59

Qy         65 VFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAST 124
              : : |   |||| ||::| |||:||| : |:|||||:  ||  ||: :  | ||||| :|
Db         60 LKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNT 118

Qy        125 PPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNGHSPI 183
              | ::|  :: : |  :     |::|    |: |:|| |  | | : | ::  : | : |
Db        119 PDNQF--HIGDASAVEIKFSNGSQDI---LLPNVIIMGAEPDLFETNSSNISLRNNYMPS 173

Qy        184 SEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGIK 243
              :  ||:  :: | |     ||:   |:          :  ||||||||||| |||||| |
Db        174 NHRFGSIAIVTFSPEYSFRFNDNCMNE---------FIQDPALTLMHELIHSLHGLYGAK 224

Qy        244 ISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKALQNF 296
                ||  ||    |  :|:          ||  |||| | ::|: :   :||   | ::
Db        225 ----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADY 278

Qy        297 QDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFT 356
              : ||::|   |  | |   :: || ::: ||   :| :| |||: :||: ::| | : ||
Db        279 KKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSFT 334

Qy        357 ETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVN 416
              | :|  :: :| | :|  :|     |   ||:::||  :||:||   |||  | |||  :|
Db        335 EFDLRTKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANLN 391

Qy        417 KEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKA-- 474
                    |:   ||   | || ::     :   || :|| :|||:|:::|:: |
Db        392 PRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPK 451

Qy        475 ---ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQ-- 529
                  :|:   | ||  ||:  :||:||: :  |   | :|     |  :|  || |
Db        452 EIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYDSNGT 505

Qy        530 SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTV 589
              | :::  |:  ::| || ||  |     |: ||:|:: ||    |:|||||:  :   |
Db        506 SDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKP 565

Qy        590 VGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFE 649
              | |:|||:|:: |: |||:|: ||||:||::|:||::|||| |||:|||  | |||:| |
Db        566 VQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALE 625

Qy        650 IGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGLIV 706
              : || ||:|| |||::|  :  ||::|::|    ||  :|  |:||||:||:|| ::|  ||
Db        626 LLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIV 685

Qy        707 SQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFNDIDFK 764
              | |:: :||||   ||:|| || ||  ||: ||| :|| |: |:|  :    |   |: :
Db        686 SNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENE 745

Qy        765 LNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLD 824
              ||| :::|:|||| |: : ||||||  :  : : ||:::|:|:|  || ||  :   | :
Db        746 LNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQHGSILGE 805

Qy        825 EVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 884
                | | |   | :|||| || || | ||| || :   | |:::|:: |:  : :|:|
Db        806 SQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTS 865

Qy        885 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 944
              || : :|:  ||          || : | : | :   |: :::||: : ||||:||| | |
Db        866 GYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNY 925

Qy        945 NNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNIS 1003
              :|  :   : ||||||:|:: |:||||||:  | ||||  |    :: : | |   : ||
Db        926 DNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTFEDNRGINQKLAFNYGNANGIS 983

Qy       1004 DYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVW 1063
              |||||| :|||||||:: :||||:|   : ||||  |: |::| ||::||:  |:::
Db        984 DYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCS-YTRYIG 1042

Qy       1064 IKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSK 1123
              |: |||| :||: ||: :||  : :|| |||||| | || :||| |    | :|
Db       1043 IRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFI----D 1098

Qy       1124 ASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISD 1183
                    |   |  :  :    || |::  |:: :||    |||:||: | :|:|
Db       1099 RRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN------ 1150

Qy       1184 ESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKT--------TYNCQIL--- 1232
                      | ||   | ||       | : |      |  |||          :|  ::
Db       1151 -------FVASK---THLF-------PLYADTATTNK--EKTIKISSSGNRFNQVVVMNS 1191

Qy       1233 ---CEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFI 1289
                  |  : |      ||       |:  ||       | ||   : :: |    || | ||
Db       1192 VGNCTMNFKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDHTNS--NGCFWNFI 1241

Qy       1290 PVDEGWTE 1297
                : || |
Db       1242 SEEHGWQE 1249


RESULT 6
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
```

```
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             32.7%;  Score 2232;  DB 1;  Length 1250;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-93;
  Matches  509;  Conservative  246;  Mismatches  468;  Indels   94;  Gaps   32;

Qy          5 IKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 64
              | :||||||:||  |: ::|    |  :||:| |:  |||:|||   |  |  |   |
Db          3 INSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPPTS 59

Qy         65 VFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAST 124
              : :  |   |||| ||::| ||||||| : |:|||||   ||: ||: :  | |||||  :|
Db         60 LKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLGNDNT 118

Qy        125 PPDKFAANVAN-VSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNGHSP 182
              |   |  | |: | |        |::     |: |:|| |  | | : | ::  : | : |
Db        119 PDGDFIINDASAVPIQ---FSNGSQSI---LLPNVIIMGAEPDLFETNSSNISLRNNYMP 172

Qy        183 ISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLYGI 242
               : |||:  :: | |       |       || |:      :  ||||||||||| ||||||
Db        173 SNHGFGSIAIVTFSPEYSFRF------KDNSM---NEFIQDPALTLMHELIHSLHGLYGA 223

Qy        243 KISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKALQN 295
              |      ||  ||     |  :|:          ||  |||| | ::|: :   :||   | :
Db        224 K----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLAD 277

Qy        296 FQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGF 355
              :: ||::|   |  | |   :: || ::: ||   :| :| |||: :||: ::| | : |
Db        278 YKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL-YSF 333

Qy        356 TETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAV 415
              || :|| :: :| | :|  :|     |   ||:::||  :||:||   |||  | |||  :
Db        334 TEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQNANL 390

Qy        416 NKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKA- 474
              |      |:   ||   |  || ::      :   || :|| :|||:|:::|:: |
Db        391 NPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTP 450

Qy        475 ----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQ- 529
                  :|:   | ||  ||:  :||:||: :  |  | :|     |  :|  || |
Db        451 KEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYDSNG 504

Qy        530 -SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANT 588
               | :::  |:  ::| || ||    |: ||:|:: ||     |:|||:  :   |
Db        505 TSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNK 564

Qy        589 VVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAF 648
              | |:||| |:: |: |||:|: |||||:||::|:||::|||| |||:|||  | |||:|
Db        565 PVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDAL 624

Qy        649 EIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGLI 705
              |: || ||:|| |||::| :  ||::|::|   ||  :|  |:||||:||:|| ::|  |
Db        625 ELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI 684

Qy        706 VSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFNDIDF 763
              || |:: :||||   ||:|| || ||  |:: ||| :|| |: |:|  :     |   |:
Db        685 VSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQIEN 744

Qy        764 KLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLL 823
              :||| :::|:|||| |: : ||||||  :  : : ||:::|:|:|  ||:||  :   |
Db        745 ELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHGSILG 804

Qy        824 DEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDS 883
              :    | | |   | :|||| || || | ||| || :   | |:::|:: |:  : :|:
Db        805 ESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDT 864

Qy        884 SGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPK 943
              ||| : :|:  ||      || : | : |    |: :::||: : ||||:|||| |
Db        865 SGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPN 924

Qy        944 YNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNI 1002
```

```
Db        925 YDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGNANGI 982

Qy       1003 SDYINKWFSITITNDRLGNANIYIYNGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFV 1062

Db        983 SDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCS-YTRYI 1041

Qy       1063 WIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS 1122

Db       1042 GIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFI---- 1097

Qy       1123 KASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNIS 1182

Db       1098 NRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN----- 1150

Qy       1183 DESYRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTF 1240

Db       1151 --------FVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNFKNN 1202

Qy       1241 GLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1203 NGNNIGLL----GFKADT----VVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249


RESULT 7
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match            32.4%;  Score 2206;  DB 1;  Length 1295;
  Best Local Similarity  38.1%;  Pred. No. 1.8e-92;
  Matches  513;  Conservative  214;  Mismatches  508;  Indels  110;  Gaps   28;

Qy          6 KNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTGV 65

Db          5 KQFNYKDPVNGVDIAYIK-IPNAGQMQPVKAFKIHNKIWVIPERDTF-TNPEEGDLNPPP 62

Qy         66 FSKDV-YEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLGNAS- 123

Db         63 EAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTI 122

Qy        124 -TPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSP 182

Db        123 DTELKVIDTNCINV------IQPDGSYRSEEL--NLVIIGPS---ADIIQFECKSFGHDV 171

Qy        183 IS---EGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239

Db        172 LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAEHRL 231

Qy        240 YGIKIS-NLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTD----MNIYNKALQ 294

Db        232 YGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNK--- 288

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354

Db        289 -FKDVASTLNKAKSIIGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI 347

Qy        355 FTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKA 414

Db        348 YTEDNFVNFFKVINRKTYLN-FDKAVFRINIVPDENYTIKDGFNLKGANLSTNFNGQNTE 406

Qy        415 VNKEAYEEIS-----LEHLVIYRIAMCKPVMYK------NTGKSEQCIIVNNEDLFFIAN 463
```

```
Db        407 INSRNFTRLKNFTGLFEFYKLLCVRGIIPFKTKSLDEGYNKALNDLCIKVNNWDLFFSPS 466

Qy        464 KDSFSKDLAKAETIAYNTQNNTIENNFSIDQL-------ILDND--------LSSGID-- 506

Db        467 EDNFTNDLDKVEEITADTNIEAAEENISLDLIQQYYLTFDFDNEPENISIENLSSDIIGQ 526

Qy        507 -LPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLN 565

Db        527 LEPMPNIERFPNG-------------KKYELDKYTMFHYLRAQEFEHGDSRIILTNSAE 572

Qy        566 DALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSII 625

Db        573 EALLKPNVAYTFFSSKYVKKINKAVEAFMFLNWAEELVYDFTDETNEVTTMDKIADITII 632

Qy        626 IPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIM 685

Db        633 VPYIGPALNIGNMLSKGEFVEAIIFTGVVAMLEFIPEYALPVFGTFAIVSYIANKVLTVQ 692

Qy        686 TISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNR 745

Db        693 TINNALSKRNEKWDEVYKYTVTNWLAKVNTQIDLIREKMKKALENQAEATKAIINYQYNQ 752

Qy        746 YSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDD 805

Db        753 YTEEEKNNINFNIDDLSSKLNESINSAMININKFLDQCSVSYLMNSMIPYAVKRLKDFDA 812

Qy        806 NLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNI 865

Db        813 SVRDVLLKYIYDNRGTLVLQVDRLKDEVNNTLSADIPFQLSKYVDNKKLLSTFTEYIKNI 872

Qy        866 SSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVV 925

Db        873 VNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLESSTIEVILKNAIV 932

Qy        926 YDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDV 985

Db        933 YNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSLNYGEIIWTLQDN 989

Qy        986 NAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSS 1045

Db        990 KQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQKPISNLGNIHAS 1049

Qy       1046 NDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQ 1105

Db       1050 NKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILKDFWGNYLQYDKP 1109

Qy       1106 YYLFNQGMQNIYIK---------YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKA 1156

Db       1110 YYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNS-----TLYEGTKFIIKKY 1164

Qy       1157 SNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVL 1216

Db       1165 ASG---NEDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVV 1220

Qy       1217 QIKKYYEKTTYN-CQILCEKDTKTFGLFGIGKFVKDYGYV-WDTYDNY--FCISQWYLRR 1272

Db       1221 VMKSKDDQGIRNKCKMNLQDNNG----------NDIGFIGFHLYDNIAKLVASNWYNRQ 1269

Qy       1273 ISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1270 VGK--ASRTFGCSWEFIPVDDGWGE 1292


RESULT 8
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
```

```
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             32.1%;  Score 2188.5;  DB 1;  Length 1274;
  Best Local Similarity   37.2%;  Pred. No. 1.1e-91;
  Matches  498;  Conservative  258;  Mismatches  477;  Indels  105;  Gaps   35;

Qy          1 MPVNIKNFNYNDPINNDDIIMME-PFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQF 59
              |||  | :||||||:|:|  |: |: |: :     |||||  |:  :||:||| | |   |  |
Db          1 MPVAINSFNYNDPVNDDTILYMQIPYEEKSK-KYYKAFEIMRNVWIIPERNTIGTNPSDF 59

Qy         60 NASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
              :    :       |||| || ||||||::||| |||| |||| |:|: ||   |  | |||
Db         60 DPPASL-KNGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGKVLLQEISYAKPYL 118

Qy        120 GNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDN-------FT 172
              ||  || |:|:      |:| |:    : :    :: ||:: | || : ::
Db        119 GNDHTPIDEFSPVTRTTSVNIKL----STNVESSMLLNLLVLGAGPDIFESCCYPVRKLI 174

Qy        173 DSMIMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHEL 232
              |  ::  : | : |||:  :: | |      ||::    ::    |  :: ||||::| |||
Db        175 DPDVV--YDPSNYGFGSINIVTFSPEYEYTFNDISGGHNS---STESFIADPAISLAHEL 229

Qy        233 IHVLHGLYG---------IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPS 283
              || ||||||         |::   |:         |     |:: ||  |||| | ::|: :
Db        230 IHALHGLYGARGVTYEETIEVKQAPL--------MIAEKPIRLEEFLTFGGQDLNIITSA 281

Qy        284 TDMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDK 343
              |||  | |:: || ||: |:||     ||: ||  :: ||   :: :| |:|:::|
Db        282 MKEKIYNNLLANYEKIATRLSEVNSAPPE-YDINEYKDYFQWKYGLDKNADGSYTVNENK 340

Qy        344 FDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKN 403
              |:::|| | : |||:|| :: :| | :|| :|   :|   |||: ||| :|||||    |
Db        341 FNEIYKKL-YSFTESDLANKFKVKCRNTYFIKY-EFLKVPNLLDDDIYTVSEGFNIG--N 396

Qy        404 LKTEFNGQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKS--EQCIIVNNEDLFFI 461
              |      ||:  :| :  : |  : ||   :  || |: :   |:    || ||| :|||:
Db        397 LAVNNRGQSIKLNPKIIDSIPDKGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNSELFFV 456

Qy        462 ANKDSFSK-DLAKAETIAYNTQ-NNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFD 519
              |:: |::: |:   : |   |  ||   ||  :|::||| :  :   : |           |
Db        457 ASESSYNENDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSQTIPQISNRTLNTLVQ-D 513

Qy        520 DIDIPVYIKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTF 577
              :  :| |     | :::  |   ::| |||||  |    |: ||:|:: ||   :|   |
Db        514 NSYVPRYDSNGTSEIEEYDVVDFNVFFYLHAQKVPEGETNISLTSSIDTALLEESK-DIF 572

Qy        578 FSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGN 637
              ||:  ::  |  | |:||::|:  || |||:|:|||||:||::|:|:|:||:| |||:
Db        573 FSSEFIDTINKPVNAALFIDWISKVIRDFTTEATQKSTVDKIADISLIVPYVGLALNIII 632

Qy        638 ETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKR 694
              |  | ||: |||: |  ||:||:||| :|::  ||::||:    ||   |  |:|:| :|
Db        633 EAEKGNFEEAFELLGVGILLEFVPELTIPVILVFTIKSYIDSYENKNKAIKAINNSLIER 692

Qy        695 DQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI 754
              : || ::|  ||| ||: :||||   ||:|| || ||  ||:  || :|| |: ::|  :
Db        693 EAKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRL 752

Qy        755 NIDF--NDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLL 812
                ::  |:|: :||: ::||: ||: |: : ||||||  :    | ||| :|:::| |||
Db        753 ESEYNINNIEEELNKKVSLAMKNIERFMTESSISYLMKLINEAKVGKLKKYDNHVKSDLL 812

Qy        813 EYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILS 872
              ||  :   | :: | |   |   | ||||:|| || | |||  ||     |  ::||
Db        813 NYILDHRSILGEQTNELSDLVTSTLNSSIPFELSSYTNDKILIIYFNRLYKKIKDSSILD 872

Qy        873 LSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDN 932
              : |   : || |||| ::: :|         || : ||  | :   |:  ::|:| : |
Db        873 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYNSRLSEVNIAQNNDIIYNSRYQN 932

Qy        933 FSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KNDSGWKVSIKGNR---IIWTLID 984
              |||:|||| ||:    |:|        |||||:|:  |:||||:|::  |   ||||| |
Db        933 FSISFWVRIPKHYKPMNHN-------REYTIINCMGNNNSGWKISLRTVRDCEIIWTLQD 985

Qy        985 VNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINS 1044
              :   ::: | |   : ||:|||||  :||||:|||: |||||:|  : | ||  |:
Db        986 TSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHV 1045

Qy       1045 SNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDT 1104
              |::| ||:: | | | :| |: | :|  ||: ||: :||  :   : ||::||| | |:
Db       1046 SDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPDPSILKNYWGNYLLYNK 1104

Qy       1105 QYYLFNQGMQNIYIKYFS-----KASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNS 1159
              :|||||   :: ||   |          | |       :  :||: || |:  ||:|
Db       1105 KYYLFNLLRKDKYITLNSGILNINQQRGVTE-----GSVFLNYK-LYEGVEVIIRKNGPI 1158

Qy       1160 RNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK 1219
                | || ||: |  |:|: :   | ||:|    |::   :  :  | : :   :: :
Db       1159 DISNTDNFVRKNDLAYINVVDRGVE-YRLYADTKSEK---EKIIRTSNLNDSLGQIIVMD 1214

Qy       1220 KYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINK 1279
                | |  |     :   || |                :  |   | ||   | |  :
Db       1215 SIGNNCTMNFQ---NNNGSNIGLLGF------------HSNNLVASSWYYNNIRRNTSS 1258

Qy       1280 LRLGCNWQFIPVDEGWTE 1297
                || |  |  : || |
Db       1259 --NGCFWSSISKENGWKE 1274


RESULT 9
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
```

```
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             30.5%;  Score 2081.5;  DB 1;  Length 1276;
  Best Local Similarity   35.1%;  Pred. No. 7.6e-87;
  Matches  471;  Conservative  250;  Mismatches  501;  Indels  121;  Gaps   28;

Qy          5 IKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 64
              :|:|||:||:|::||: :     :    |  ||| |   ||::||||:    |
Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPS------- 57

Qy         65 VFSK------DVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPY 118
                ||          ||||:|| || :|| ||| :|||| ||| :  |::|:: :|   |:
Db         58 -LSKPPRPTSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPF 116

Qy        119 LGNASTPPD--KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMI 176
              :|::||| |   |  :  |::: |   : |:      :  :::|||| | : | :| |:
Db        117 MGDSSTPEDTFDFTRHTTNIAVEK--FENGSWKVTNIITPSVLIFGPLPNILD-YTASLT 173

Qy        177 MNGH--SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234
              : |   :|  ||||   :::  |  |  |::|  |: :::  :  :  || : |||| |
Db        174 LQGQQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTH 233

Qy        235 VLHGLYGIKI-SNLPITPNTKE-FFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKA 292
              || |||| | |:  | |   | || |   || |||||||  |  :|        :  ||
Db        234 SLHQLYGINIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKA 293

Qy        293 LQNFQDIANRLNIVSSAQGSG--IDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKA 350
              | :::||| ||| ::    |    :|  ||:|:  ||:| :|  | : |: |||: ||
Db        294 LGHYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSD 353

Qy        351 LMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNG 410
              |    :|   : :| :| |  ||| :  |:    :||: |||  :|||: :|    | :|
Db        354 LTNVMSEVVYSSQYNVKNRTHYFSRHYLPV-FANILDDNIYTIRDGFNLTNKGFNIENSG 412

Qy        411 QNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKD 470
              ||   |  | :::| | :|     :|  : ||:      || | |  | ::|:||| |::
Db        413 QNIERN-PALQKLSSESVVDLFTKVCLR-LTKNSRDDSTCIKVKNNRLPYVADKDSISQE 470

Qy        471 LAKAETIAYNTQNNTIENNFSIDQLILDNDLSSG---ID--LPNENTEP-------FTNF 518
              : : :  |   |      : ||:|: |||  :      :|  ||| | ||          :
Db        471 IFENKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFY 530

Qy        519 DDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFF 578
              |||               :||  : : || :|   :|:||: || |: :||  :||:|||
Db        531 DDI------------TKYVDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFL 578

Qy        579 STNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNE 638
              :| || |  | | ||:||   |::|||:   :| |:||':||||:||||||||||:||
Db        579 -PSLAEKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNS 637

Qy        639 TAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKW 698
              : ||  ||   | | |:|  ||  :| :| ||  | :  :  || || | |::| ::|
Db        638 ALRGNFNQAFATAGVAFLLEGFPEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRW 697

Qy        699 TDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDF 758
              | |  :|| ||| : |||  |  :||::|: |: ||:  |: :| :||  || ||
Db        698 KDSYQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQV 757

Qy        759 NDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTN 818
              ::   |:  |:|:|||: || :||::||   |:|  : :|  ||   | :|:  ||::
Db        758 ENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSH 817

Qy        819 ELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGG 878
              : |: ||: ||:|||   ::::||::  || :::|  : | | ::|: : ||||  :
Db        818 NIILVGEVDRLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKN 877

Qy        879 RLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFW 938
              |:|:||| | : || :|  | |    |||::|    |  : :  ::| ::::| |::||
Db        878 ALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSSS-GDKIIVNLNNNILYSAIYENSSVSFW 936
```

```
Qy        939 VRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSI 998
              ::  |   |       ||||||: |: :||||: |:   | | | ||| | ||: |:||
Db        937 IKISKDLTNS-----HNEYTIINSIEQNSGWKLCIRNGNIEWILQDVNRKYKSLIFDYSE 991

Qy        999 KDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDT 1058
                : :  | |||| :||||: :|   :|||| ||:|:|| :||  :     | |  :    |
Db        992 SLSHTGYTNKWFFVTITNNIMGYMKLYINGELKQSQKIEDLDEVKLDKTIVFGIDENIDE 1051

Qy       1059 TKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYI 1118
               : :||:||||| :||:  ::: :|  |    | :||:|||||::||:||: |    : ||
Db       1052 NQMLWIRDFNIFSKELSNEDININVYEGQILRNVIKDYWGNPLKFDTEYYIINDNYIDRYI 1111

Qy       1119 KYFSKASMGETAPRTNFNNAAINY---QNLYLGLRFIIKKASN----SRNINNDNIVREG 1171
                        || :|     : |     || |    ||  |:    || :| |||:
Db       1112 -----------APESNV-LVLVQYPDRSKLYTGNPITIKSVSDKNPYSRILNGDNII--- 1156

Qy       1172 DYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQI 1231
                           :::| ||::         | |  | |   |  :  :   | ::
Db       1157 ---------------LHMLYNSRKYMI------IRDTDTIY-ATQGGECSQNCVYALKL 1193

Qy       1232 LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--------------FCISQWYLRRISEN 1276
                    |:| |   |    |    : ::                 |     |      |
Db       1194 QSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTN 1253

Qy       1277 INKLRLGCN--WQFIPVDEGWTE 1297
                  |  :  |:||  | || |
Db       1254 YETKLLSTSSFWKFISRDPGWVE 1276
```

RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO   STANDARD;   PRT; 1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC   RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC   AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC   WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC   INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC   ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC   neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC   detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC   heavy chain (H). The light chain has the pharmacological activity,
CC   while the N- and C-terminal of the heavy chain mediate channel
CC   formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC   botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC   STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC   BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET   0   0
FT   CHAIN   1   448   BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN   449   1290   BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL   228   228   ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   229   229   BY SIMILARITY.
FT   METAL   232   232   ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID   436   452   INTERCHAIN (PROBABLE).
FT   CONFLICT   84   84   P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA; 148734 MW; 71FBE379F97129E8 CRC64;

  Query Match   27.9%; Score 1903; DB 1; Length 1290;
  Best Local Similarity   34.1%; Pred. No. 8.9e-79;
  Matches 463; Conservative 240; Mismatches 523; Indels 130; Gaps   32;

```
Qy          2 PVNIKNFNYNDPINNDDIIMMEPF-----NDPGPGTYYKAFRIIDRIWIVPERFTYGFQP 56
              |: | ||||:||::| :|: ::       |:|      |||||   ||::|:||:    |
Db          1 PITINNFNYSDPVDNKNILYLDTHLNTLANEP-----EKAFRITGNIWVIPDRFSRNSNP 55

Qy         57 DQFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAI 116
                :  |    | |     |||| || ||::|| ||| :|||| ||||:  |: |:  :   |
Db         56 N-LNKPPRVTSPK-SGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDI 113

Qy        117 PYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSM- 175
              |: || :|| : |  :|   |:: |  |     :   :  ::|| ||   : |  | :
```

```
Db        114 PFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFK 173

Qy        176 IMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHV 235

Db        174 LTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHA 233

Qy        236 LHGLYGIKISN----LPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNK 291

Db        234 MHNLYGIAIPNDQTISSVTSNI--FYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEK 291

Qy        292 ALQNFQDIANRLNIVSSAQGSGID--ISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYK 349

Db        292 ALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYN 351

Qy        350 ALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFN 409

Db        352 ELTQIFTEFNYAKIYNVQNRKIYLSNVYTPV-TANILDDNVYDIQNGFNIPKSNLNVLFM 410

Qy        410 GQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNT--GKSEQC--IIVNNEDLFFIANKD 465

Db        411 GQNLSRN-PALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDIS 469

Qy        466 SFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSS---------GID-----LPNEN 511

Db        470 DVKTDIFLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGEN 529

Qy        512 TEPFTNFDDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNN 571

Db        530 QVFYDN--------------RTQNVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNS 574

Qy        572 NKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGP 631

Db        575 AKVYTYFPT-LANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGP 633

Qy        632 ALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNAL 691

Db        634 ALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCL 693

Qy        692 KKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDK 751

Db        694 EQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDK 753

Qy        752 MNINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDL 811

Db        754 ENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKL 813

Qy        812 LEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAIL 871

Db        814 INLIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKIL 873

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNS--ENSNITAHQSKFVVYDSM 929

Db        874 SLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSM 933

Qy        930 FDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKS 989

Db        934 YESFSISFWIRINKWVSN------LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSE 987

Qy        990 KSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDID 1049

Db        988 QSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYINGKLIDTIKVKELTGINFSKTIT 1046

Qy       1050 FKLINCTDTTKF--------VWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLR 1101

Db       1047 FEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLQYTNVVKDYWGNDLR 1106

Qy       1102 YDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAAINYQNLYLGLRFIIKKASNSR 1160

Db       1107 YNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE----------GYKIIIKRI---R 1153

Qy       1161 NINNDNIVREGDYIYLNI-----------------DNISDES-YRVYVLVNSKEIQTQL 1201

Db       1154 GNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 1213

Qy       1202 F--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTY 1259

Db       1214 IFQIQPMNN--TYYYASQI----FKSNFNGENIS-------GICSIGTYRFRLGGDWYRH 1260

Qy       1260 DNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 1293

Db       1261 -NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1290


RESULT 11
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match             10.0%;  Score 680.5;  DB 1;  Length 1162;
  Best Local Similarity   22.3%;  Pred. No. 1.3e-23;
  Matches  307;  Conservative  232;  Mismatches  520;  Indels  317;  Gaps   55;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYG----FQP 56

Db          1 MKIN-GNLNIDSPVDNKNVAIVRSRNQ----MFFKAFQVAPNIWIAPERY-YGESLKINE 54

Qy         57 DQFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAI 116

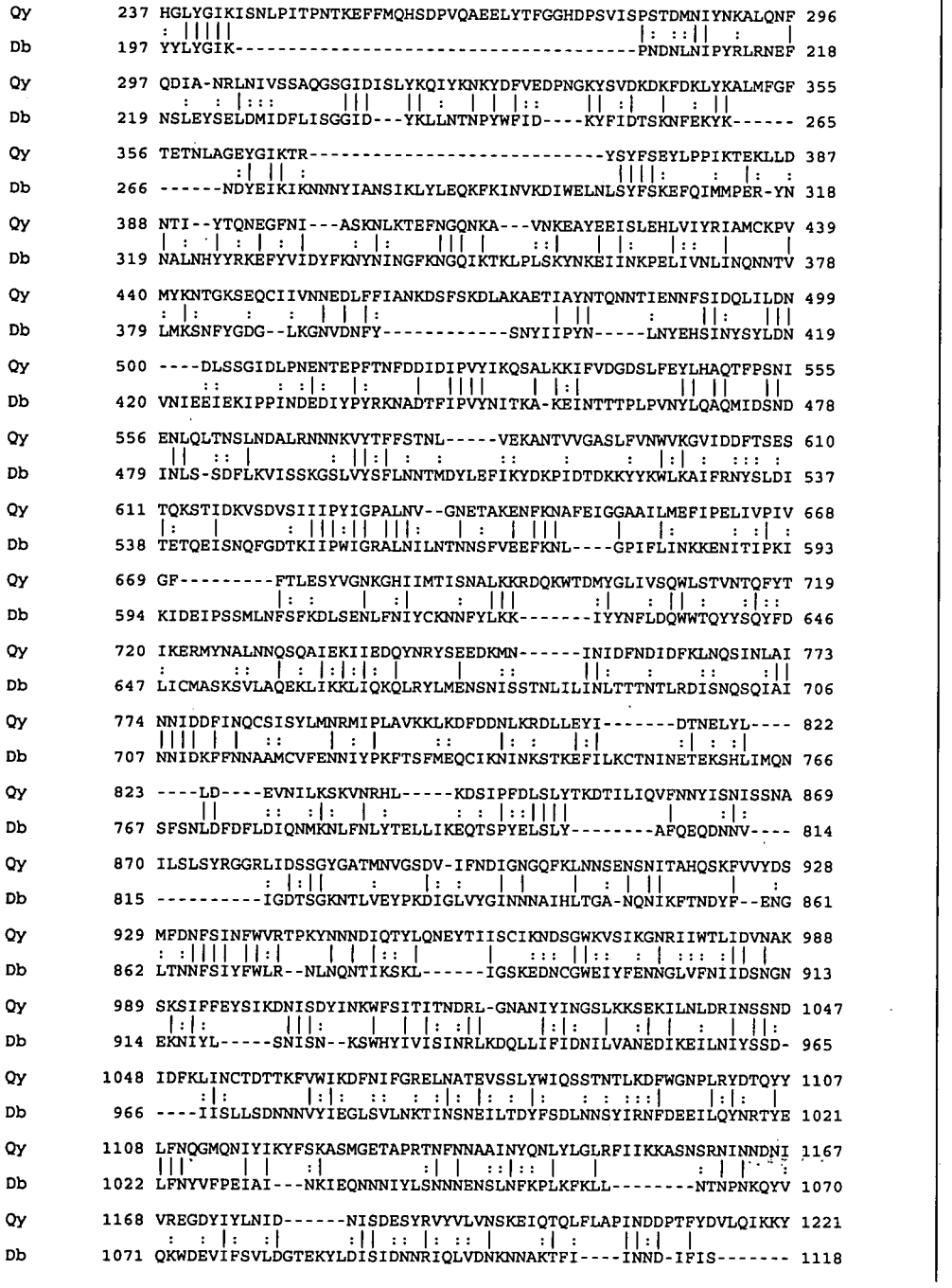
Db         55 DQ-KSDGGI--------YDSNFLSTNNEKDEFLQATIKLLQRINNNVVGAKLLSLISTAI 105

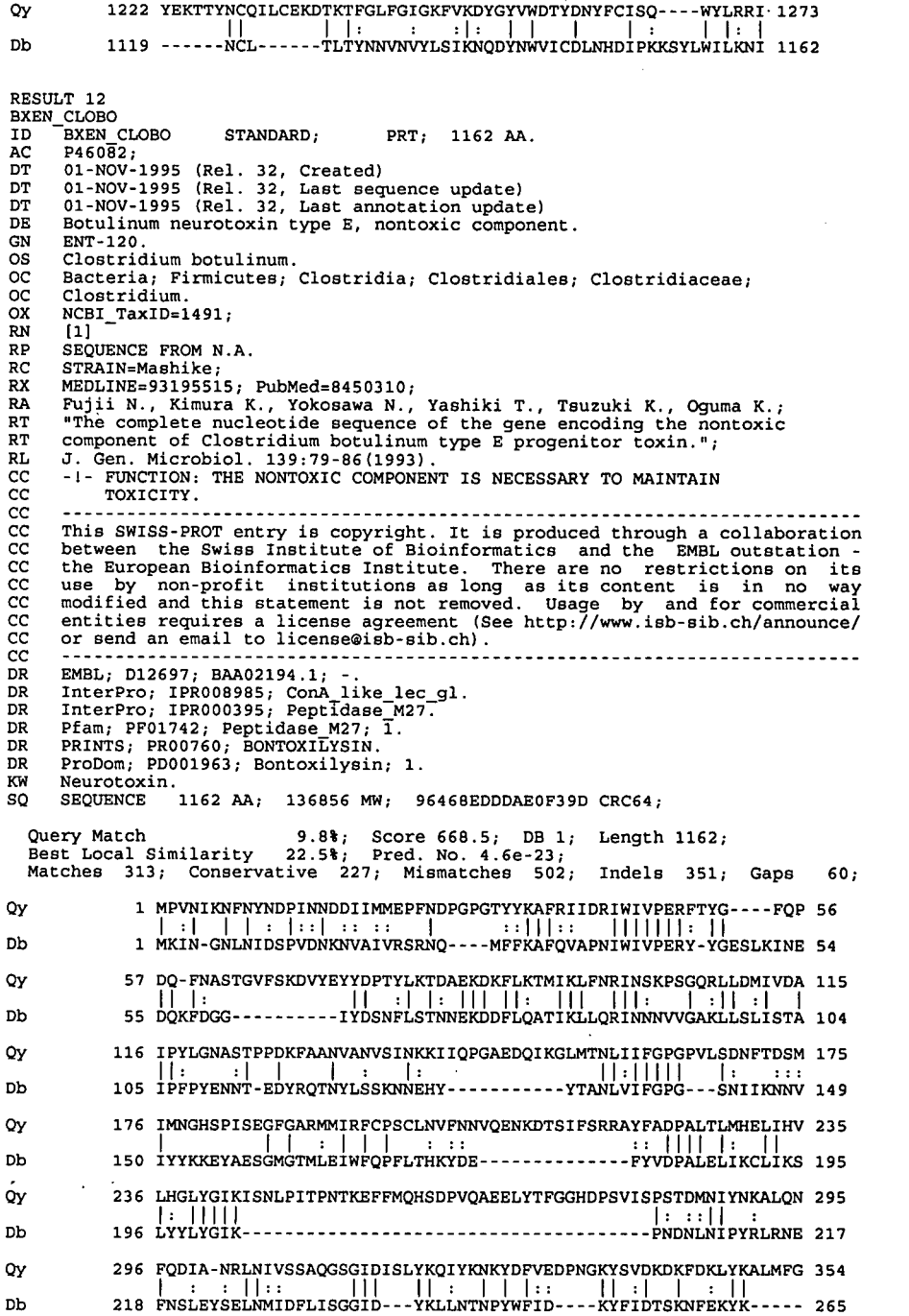
Qy        117 PYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMI 176

Db        106 PFPYENNT-EDYRQTNYLSSKNNEHY----------YTANLVIFGPG---SNIIKNNVI 150

Qy        177 MNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVL 236

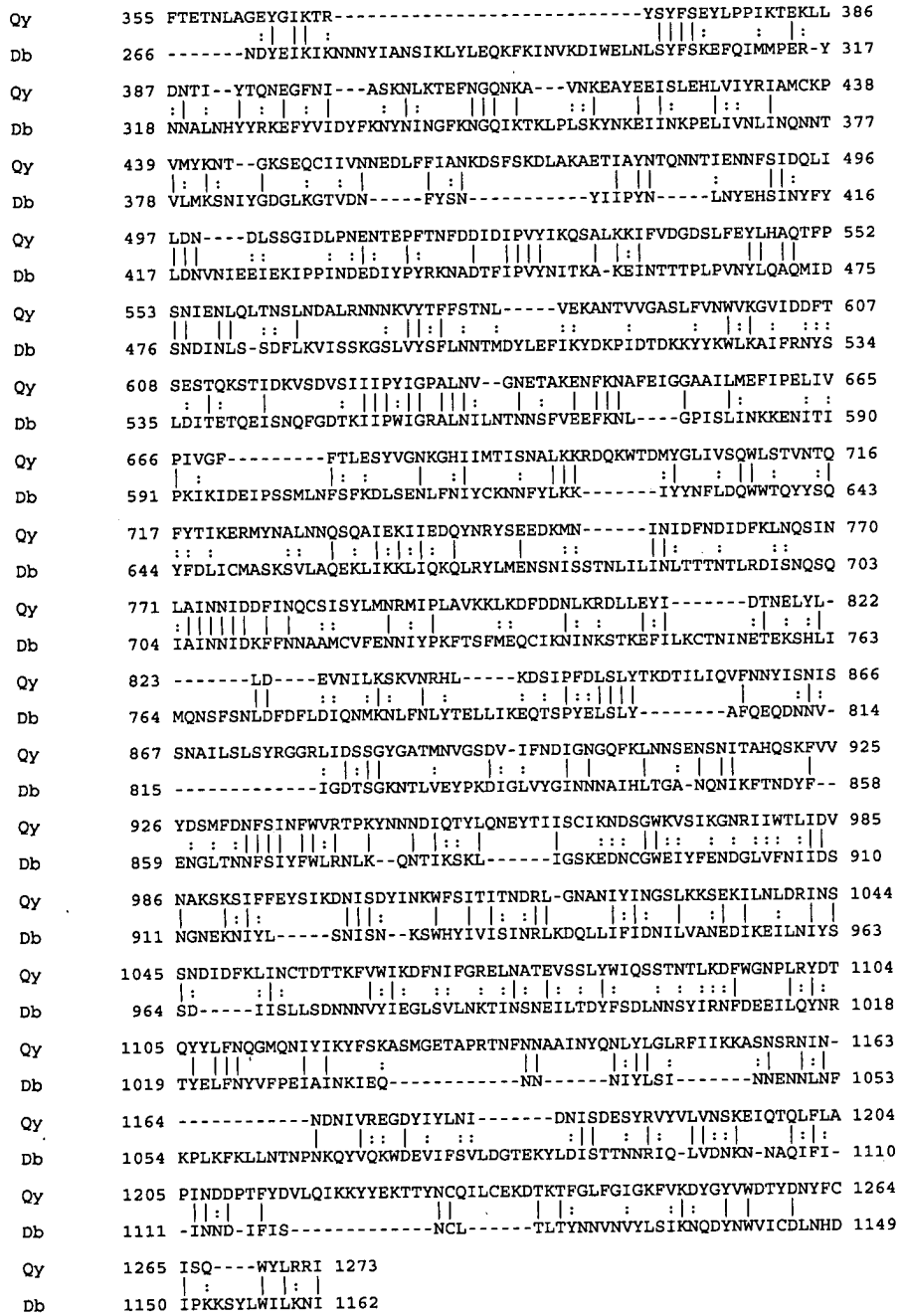
Db        151 YYKKEYAENGMGTMLEIWFQPFLTHKYDE-------------FYVDPALELIKCLIKSL 196
```





RESULT 12
BXEN_CLOBO
ID BXEN_CLOBO STANDARD; PRT; 1162 AA.
AC P46082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
GN ENT-120.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mashike;
RX MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
RT component of Clostridium botulinum type E progenitor toxin.";
RL J. Gen. Microbiol. 139:79-86(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D12697; BAA02194.1; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000395; Peptidase_M27.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDDAE0F39D CRC64;

Query Match 9.8%; Score 668.5; DB 1; Length 1162;
Best Local Similarity 22.5%; Pred. No. 4.6e-23;
Matches 313; Conservative 227; Mismatches 502; Indels 351; Gaps 60;



```
RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO      STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             9.1%;  Score 619;  DB 1;  Length 1196;
  Best Local Similarity   22.3%;  Pred. No. 8.2e-21;
  Matches  308;  Conservative  250;  Mismatches  480;  Indels  346;  Gaps   70;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQP---- 56
              | :|  : | | |::| :::::        |::|||::   ||: |||: || :|
Db          1 MDIN-DDLNINSPVDNKNVVIVRARK---TNTFFKAFKVAPNIWVAPERY-YG-EPLDIA 54

Qy         57 DQFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAI 116
              :::    |:          ||  :|  |:|:: ||: :| |  |||:  ||::|| :|  ||
Db         55 EEYKLDGGI--------YDSNFLSQDSERENFLQAIIILLKRINNTISGKQLLSLISTAI 106

Qy        117 P----YLGNASTPPDKFAANVANVSINKKI--------------------IQPGAEDQIK 152
              |    |:|   : |: |       | |||:                    |:  :::
Db        107 PFPYGYIGGGYSSPNIFTFGKTPKS-NKKLNSLVTSTIPFPFGGYRETNYIE--SQNNKN 163

Qy        153 GLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDT 212
                :|:|||||  : :|   ::|    :    | |    | | |      :|
Db        164 FYASNIIIFGPGSNIVEN---NVIYYKKNDAENGMGTMAEIVFQPLLTYKYNK------- 213

Qy        213 SIFSRRAYFADPALTLMHELIHVLHGLYGIKISNLPITP-------NTKEFFMQHSDPVQ 265
                  :: |||: |   ||  |: ||||| |:  : |        : |:|      :
Db        214 -------FYIDPAMELTKCLIKSLYFLYGIKPSDNLVVPYRLRTELDNKQF-----SQLN 261

Qy        266 AEELYTFGGHDPSVI--SPSTDMNIY----NKALQNFQDIANRLNIVSSAQGSGIDISLY 319
                :|   || |   |  :|    | |      | : :::|       ::| |: | : |
Db        262 IIDLLISGGVDLEFINTNPYWFTNSYFPNSIKMFEKYKNIYKTEIEGNNAIGNDIKLRLK 321

Qy        320 KQIYKNKYDFVEDPNGKY------SVDKDKFDKLYKALMFGFTETNLAGEYGIKTRYSYF 373
                ::   |  | : : |  |      |:  |:|   |                : : :| |
Db        322 QKFQINVQD-IWNLNLNYFCQSFNSIIPDRFSNALK-------------HFYRKQY-YT 365
```

```
Qy        374 SEYLPPIKTEKLLDNTIYTQ----NEGFNIASKNLKTEFNGQNKAVNKEAYEEISLEHLV 429
Db        366 MDYTDNYNINGFVNGQINTKLPLSNKNTNIISK--------PEKVVNLVNENNISLMKSN 417

Qy        430 IYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENN 489
Db        418 IYGDGL------KGT----------TED-FYSTYKIPYNEEY----EYRFNDSDNFPLNN 456

Qy        490 FSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGDSL-FEYLHA 548
Db        457 ISIEEV---DSIPEIIDI-----NPYK--DNSDNLVFTQITSMTEEVTTHTALSINYLQA 506

Qy        549 QTFPSNIENLQLTNSLNDALRNNNK--VYTFFSTNLVEKANTVVG------ASLFVNWVK 600
Db        507 QI--TNNENFTLSSDFSKVVSSKDKSLVYSFLD-NLMSYLETIKNDGPIDTDKKYYLWLK 563

Qy        601 GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI 660
Db        564 EVFKNYSFDINLTQEIDSMCGINEVVLWFGKALNILN--TSNSFVEEYQDSGAISLISKK 621

Qy        661 PELIVPIVGFFTL-ESYVG-------NKGHIIMTISNALKKR------DQKWTDMYGLIV 706
Db        622 DNLREPNIEIDDISDSLLGLSFKDLNNKLYEIYSKNIVYFKKIYFSFLDQWWTEYY---- 677

Qy        707 SQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINID--------- 757
Db        678 --------SQYFELICMAKQSILAQESLVKQIVQ---NKFTDLSKASIPPDTLKLIRETT 726

Qy        758 ---FNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEY 814
Db        727 EKTFIDLS---NES-QISMNRVDNFLNKASICVFVEDIYPKFISYMEKYINNINIKTREF 782

Qy        815 I-------DTNELYLLDEV----------------NILKSKVNRHLKDSI-PFDLSLYTK 850
Db        783 IQRCTNINDNEKSILINSYTFKTIDFKFLDIQSIKNFFNSQVEQVMKEILSPYQLLLFAS 842

Qy        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910
Db        843 K---GPNSNIIEDISGKNTLIQYTESIELV----YGV--------------NGESLYLK 880

Qy        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970
Db        881 SPNETI-KFSNKFFT-NGLTNNFTICFWLRFT--GKNDDKTRL------IGNKVNNCGWE 930

Qy        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL-GNANIYINGS 1029
Db        931 IYFEDNGLVFEIIDSNGNQESVYLSNIINDN-------WYYISISVDRLKDQLLIFINDK 983

Qy       1030 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 1089
Db        984 NVANVSIDQILSIYSTN-----IISLVNKNNSIYVEELSVLDNPITSEEVIRNYFSYLDN 1038

Qy       1090 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 1149
Db       1039 SYIRDSSKSLLEYNKNYQLYN------YV--FPETSLYE----VNDNNKS------YLSL 1080

Qy       1150 RFIIKKASNSRNINNDNIVREGDYIYLNID----------------------NISDES 1185
Db       1081 K-------NTDGINISSV----KFKLINIDESKVYVQKWDECIICVLDGTEKYLDISPEN 1129

Qy       1186 YRVYVLVNSKE------IQTQLFLAPINDDPTF-YDVLQIKKYYEKT----TYNCQILCE 1234
Db       1130 NRIQ-LVSSKDNAKKITVNTDLFRP---DCITFSYN----DKYFSLSLRDGDYNWMI-CN 1180

Qy       1235 KDTK 1238
Db       1181 DNNK 1184


RESULT 14
TOXA_CLODI
ID   TOXA_CLODI     STANDARD;      PRT;  2710 AA.
AC   P16154;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Toxin A.
GN   TOXA OR TCDA.
OS   Clostridium difficile.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1496;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90221894; PubMed=2109310;
RA   Sauerborn M., von Eichel-Streiber C.;
RT   "Nucleotide sequence of Clostridium difficile toxin A.";
RL   Nucleic Acids Res. 18:1629-1630(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90129305; PubMed=2105276;
RA   Dove C.H., Wang S.Z., Price S.B., Phelphs C.J., Lyerly D.M.,
RA   Wilkins T.W., Johnson J.L.;
RT   "Molecular characterization of the Clostridium difficile toxin A
RT   gene.";
RL   Infect. Immun. 58:480-488(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RA   von Eichel-Streiber C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Only after the enteral delivery of the enterotoxin A may
CC       the characteristic disease called pseudomembranous colitis be
CC       induced.
CC   -!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC       REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC       DIFFERENT OLIGOPEPTIDES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51797; CAA36094.1; -.
DR   EMBL; M30307; AAA23283.1; -.
DR   EMBL; X92982; CAA63564.1; -.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR007577; Gly_transf_sug.
DR   Pfam; PF01473; CW_binding_1; 28.
DR   Pfam; PF04488; Gly_transf_sug; 1.
KW   Toxin; Enterotoxin.
SQ   SEQUENCE   2710 AA;  308052 MW;  0A6E52CE84C14421 CRC64;

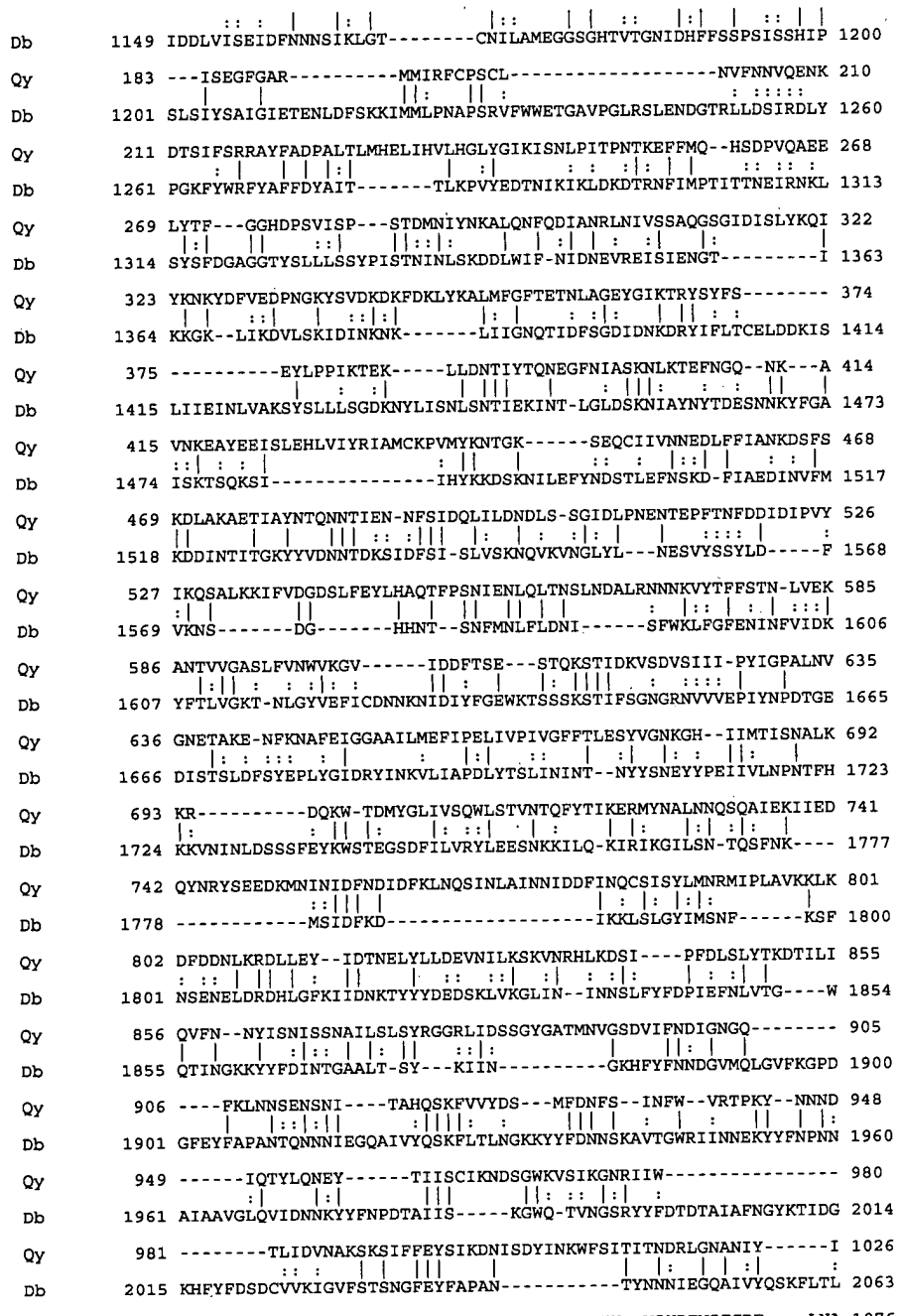
  Query Match             4.0%;  Score 273.5;  DB 1;  Length 2710;
  Best Local Similarity   20.0%;  Pred. No. 9.2e-05;
  Matches  321;  Conservative  240;  Mismatches  522;  Indels  519;  Gaps   99;

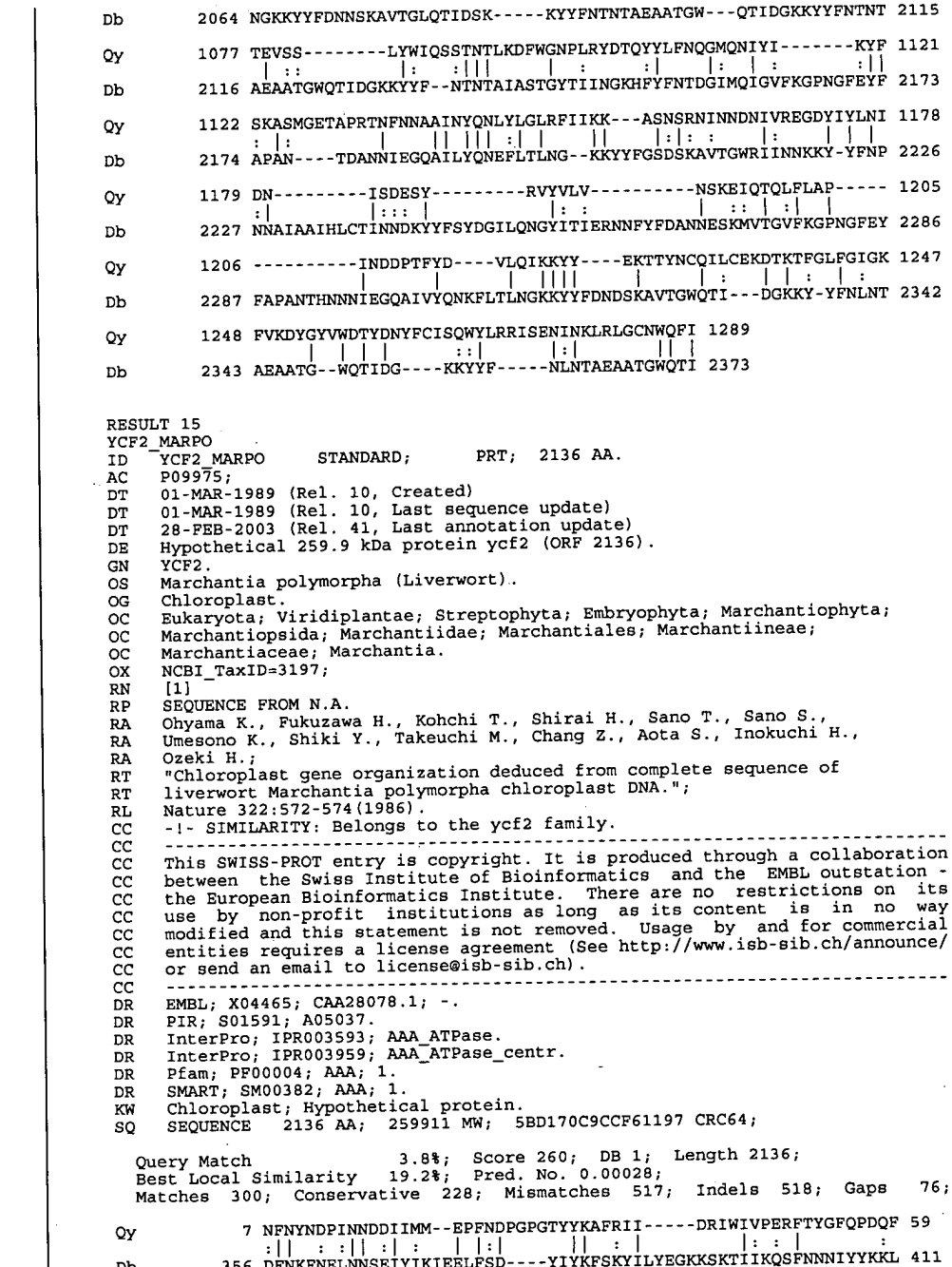
Qy          7 NFNYNDPINND-----DIIMMEPFNDPGPGTYYKAFRII--------DRIWIVPERFTYG 53
Db        972 DYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAVNDTINVLP-TITEG 1030

Qy         54 FQ-----PDQFNASTGVFSKDVYEYYDPTYLK-------------------------TD 82
Db       1031 IPIVSTILDGINLGAAI--KELLDEHDPLLKKELEAKVGVLAINMSLSIAATVASIVGIG 1088

Qy         83 AEKDKFLKTMIKLFNRINSKPSGQRLL-DMIVDAIPYLGNAS---------TPPDKFAAN 132
Db       1089 AEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSESKKYGPLKTEDDKILVP 1148

Qy        133 -----VANVSINKKIIQPGAEDQIKGLMTNLIIF--GPGPVLS---DNFTDSMIMNGHSP 182
```





RESULT 15
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT; 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC -!- SIMILARITY: Belongs to the ycf2 family.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; X04465; CAA28078.1; -.
DR PIR; S01591; A05037.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_centr.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 3.8%; Score 260; DB 1; Length 2136;
Best Local Similarity 19.2%; Pred. No. 0.00028;
Matches 300; Conservative 228; Mismatches 517; Indels 518; Gaps 76;

```
Qy        60 NASTGVFSKDVYEYYDPTYLKTDAEKDK-------FLKT----------MIKLFNRINSK 102
             |:   :|: :   |:|   |  |  |         |||:           |  | : |||
Db       412 NS---IFNFNTIFYFDSNNLLFDWLKKNYYINNKPFLKSFLIYSSISNQFILFFKQKNSK 468

Qy       103 PSGQRLL-----DMIVDAIP-----YLGNASTP--------------PDKFAAN-VANVS 137
               : |:     |:| :         : | |                :||  | ::  :
Db       469 SFNKNLVKKNSKDVITNVFSKENKIEINNFSKSIYYAFFEILSINEIDNKFVINKISLKN 528

Qy       138 INKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGHSPISEGFGARMMIRFCP 197
             |||| :   ::||                |||| :   :  |: | :          |
Db       529 INKKKQKRFYLNKIKS-------------SDNFRFINLWKIKNYSSQQFVSNNSFLLNP 574

Qy       198 SCLNVFNNVQEN------------KDTSIFSRRAYF----------------------A 222
             :   |  :|:|             |   :||  ||
Db       575 A----FEILQQNYYLKKKNILFFKKLNEVFSNFFYFQYYKCKKLNIFLKFASLEKILKKR 630

Qy       223 DPALTLMHELIHVLH-------GLYGI-------------KISNLPITPNTK--EFFMQ 259
             :   |:  :|    :         | | |               :  |    || |   |:
Db       631 NKKFTISIKLFKKFYKNKLNENGEYKIESQILQNEKELNKKRKKNFQFNPNIKILSFYNS 690

Qy       260 HSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLY 319
               :  : | |             : |: |  |  :: |:|:| ::|:::          |
Db       691 SKKNIYLQNKYFF-----------NKNLINNKLITWKKISNKL-VISNSE--------Y 729

Qy       320 KQIYKNK-----YDFVEDP-------NGK-YSVDKDKFDKLYKALMFGFTETNLAGEYGI 366
             :| ||     : | ::          | | :::    ||||  |  :   | |         |
Db       730 NKIIWNKKNMKFFSFSKNSVLDTFFFNKKSFNIITVIFDKL-KKIQLNFQE--------I 780

Qy       367 KTRYSYFSEYL--PPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYE--- 421
             :   : || :     ||  |:  |: :          ::|| | |:  :|  |     |
Db       781 QKILNCFSLFFNSKNIKKTKIFKNSYF--------INENLTTTFSFNDKEFNIFFLELFI 832

Qy       422 -EISLEHL-------VIYRIAMCKPVMYKNTGKSEQCIIVNNEDL---FFIANKDSFSKD 470
              ||: : |       : |||   | :::          :  |  |   ||   |     |
Db       833 SEINNDFLMRFFKKYLYYRIYKDKEILFNP---------IENRQLLQNFFEKTKILTFID 883

Qy       471 LAKAETIAYNT------QNNTIENN----------FSIDQ--LILDNDLSSGIDLPN--- 509
               :   : ||        :  ||:||          |  |:   :| |:: | |:  |
Db       884 FLQDPELNYNNRFIFHLEKKTIKNNNLLYLRLLKIFLKDKRNFLLINEIKSFIEKKNNLF 943

Qy       510 -----------ENTEPFTNFDDIDIPVYIKQSALK-KIFVDGDSLFE-YLHAQTFPSNIE 556
                       :|:  |  ||:|    ::||      :| ::  : ||  |  :|:  ::
Db       944 IKSQLSNVLLVKNSYKF--FDNIFNFHFLKQKEKNIEIILNNQNYFEKSLLKKTY---LK 998

Qy       557 NLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTI 616
             || | || :   : : |:: |   |:: | |                |:  :|  | |::
Db       999 NLNLNNSYS---KFSYKIFIFQLLNILNKNNYKT-----FQWISELI--FYSKNLNYKIQ 1048

Qy       617 DKVSDVSII----IPYIGPALNVGNETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFT 672
             :|:   :      | |    :   |   | |    |:   :                 |||
Db      1049 NKIEKNNYCYNKNISYKKKKIKTVNFFEKNNL---FQTNNS--------------WFFT 1090

Qy       673 LESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQS 732
             || :  |  :|:: |   :::   : ||:      : :   | :|: :| :
Db      1091 LEWWEYNT-YILLQI---IQETFFQITDVLEYFKKKKIIEKNLKFF-LKSK--------- 1136

Qy       733 QAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFINQCS-------- 784
               ||     : :: : | |:   ||:|::| |  :|   ::  : || |:
Db      1137 ----KISLKTLSFHNFKLKWNLRF-FNEINYKKNYLLNFLWSDF-NLINNCNNLYWVIFS 1190

Qy       785 ------------------------------ISYLMNRMIPL-----------AVKKLKDF 803
                                           | || :|   |           |: | ::
Db      1191 LVIFIFLYYQKIFSIIIGSDCFHLWKNFEIIQYLTDRSRSLYFTKLTRRNKTALNKTENL 1250

Qy       804 DDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYIS 863
               ::|  ||   : ||| : |: |  :|     |       |  :  |      .||
Db      1251 LSYFFQNLTHYITNIKFYLLTKKNLKKWLINNKTLDLSRRKRKLLVQSLITHNKIQNYGF 1310

Qy       864 NISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKF 923
               ::||     || | :: :  |         : |  :   ||   |  |:::    :|:
Db      1311 ELNSNKQFFTSYFGYQITNQQGL---------LYFQYL--AQFFQKNLINNSLDL-ANKW 1358

Qy       924 VVYDSMFDNFSINFWVRTPKYNNNDIQTYLQN---------EYT--IISCIKNDSGWKVS 972
```

```
                :|:      ||     :|     |:|:   ||          |:  |:     ::|
Db      1359 IVFSFWHKIFSSQKLRQT-----NNIELGFQNIPVPLQFGLSYSKGILLIGPIETGRSYL 1413

Qy       973 IKGNRIIWTLIDVNAKSKSIFFEYSIKD---NISDYINK-WFSITITNDRLGNANIYING 1028
             ||          :: |:|    |: ||     |  | | : | :| |
Db      1414 IK---------NLAAESYVPLFKISINKLLYNKPDVITESWMNILIE------------ 1451

Qy      1029 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 1088
                    :| |:| : |   |:  |      :||:: :    :||   ::
Db      1452 ---------SLRRLNLTLDFAKKMSPC-----IIWIQNIH----QLNVNRLT-------- 1485

Qy      1089 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 1148
                       :  | : |       | :|||          :|:|:   |  : :|
Db      1486 ---------QNVESDPTFLL------GILLKYF----------QTDFSKTKKN-NIIVIG 1519

Qy      1149 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQ---------- 1198
                : ||   :    |    |    | : : |||    :  :|:  |  |
Db      1520 STHLPKKVDPALISPN----RLDKIINVRLFNISQRKKQFPLLLKKKNFQLKENLFFLNE 1575

Qy      1199 -----------------TQLFLAPINDDPTFYDVLQIKKYYEKTTYNC------------ 1229
                             :: |  |  : :| |   :|  : :  :
Db      1576 FGSRTMGYNLRDLSALTNEVLLISITKNRSFIDTDTLKLAFHRQIFGLTYTNNKLNFDRI 1635

Qy      1230 -QILCEKDTKT-----------FGLFGIGKFVKDYGYVWDTYDNYFCISQWYLR-RISEN 1276
              :|:  |  ||             |  || |      :|    |:: :|:|||   | |:
Db      1636 FKIVIYKVGKTIIQNILIKSSSMNLLNIGNF------LWK--KNFYYLSKWYLEPSIDES 1687

Qy      1277 INK 1279
               | |
Db      1688 IIK 1690
```

Search completed: September 10, 2004, 03:36:02
Job time : 24.8251 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 59.3442 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-77
Perfect score: 6816
Sequence: 1 MPVNIKNFNYNDPINNDDII.....NKLRLGCNWQFIPVDEGWTE 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4007	58.8	1291	2	Q9ZAJ8	Q9zaj8 clostridium
2	4006	58.8	1291	2	Q933K0	Q933k0 clostridium
3	4001	58.7	1291	2	Q8GR96	Q8gr96 clostridium
4	3996	58.6	1291	2	Q93G71	Q93g71 clostridium
5	3932	57.7	1291	2	Q08077	Q08077 clostridium
6	2488.5	36.5	1310	2	Q93N27	Q93n27 clostridium
7	2328.5	34.2	1268	2	Q45851	Q45851 clostridium
8	2308	33.9	1278	2	Q57236	Q57236 clostridium
9	2259.5	33.1	1251	2	Q9K395	Q9k395 clostridium
10	2234	32.8	1252	2	Q8KZM3	Q8kzm3 clostridium
11	2234	32.8	1255	2	Q9FAR6	Q9far6 clostridium
12	2230.5	32.7	1280	2	Q9ZAJ5	Q9zaj5 clostridium
13	2077	30.5	1275	12	Q9QTG7	Q9qtg7 clostridium
14	2028	29.8	1285	2	Q45967	Q45967 clostridium
15	2023	29.7	1285	2	Q9LBR1	Q9lbr1 clostridium
16	1963.5	28.8	1280	2	Q45849	Q45849 clostridium
17	1957.5	28.7	1280	2	Q9LBS7	Q9lbs7 clostridium
18	1952.5	28.6	1280	2	Q841S3	Q841s3 clostridium
19	1908	28.0	1291	2	Q93HT3	Q93ht3 clostridium
20	1477	21.7	451	2	Q9R631	Q9r631 clostridium
21	1212	17.8	361	2	Q45846	Q45846 clostridium
22	1207	17.7	361	2	Q45848	Q45848 clostridium
23	1164.5	17.1	441	2	Q9X708	Q9x708 clostridium
24	800.5	11.7	367	2	Q45862	Q45862 clostridium
25	791.5	11.6	367	2	Q45861	Q45861 clostridium
26	674	9.9	1198	2	Q45893	Q45893 clostridium
27	672	9.9	1198	2	O69277	O69277 clostridium
28	669	9.8	1198	2	O06018	O06018 clostridium
29	667.5	9.8	1197	2	P71117	P71117 clostridium
30	667.5	9.8	1197	2	O33871	O33871 clostridium
31	666.5	9.8	1160	2	O87710	O87710 clostridium
32	665	9.8	1197	2	Q45888	Q45888 clostridium
33	656	9.6	1197	2	Q9ZAJ9	Q9zaj9 clostridium
34	641	9.4	1193	2	Q45880	Q45880 clostridium
35	640	9.4	1193	2	P71107	P71107 clostridium
36	640	9.4	1193	2	Q45914	Q45914 clostridium
37	633	9.3	1163	2	Q45850	Q45850 clostridium
38	632.5	9.3	451	2	Q9LA13	Q9la13 clostridium
39	631.5	9.3	1162	2	Q9ZAJ6	Q9zaj6 clostridium
40	629	9.2	1165	2	Q45844	Q45844 clostridium
41	625	9.2	1165	2	Q45887	Q45887 clostridium
42	619	9.1	347	2	Q840G4	Q840g4 clostridium
43	618	9.1	1196	2	Q45916	Q45916 clostridium
44	618	9.1	1196	2	Q93HT4	Q93ht4 clostridium
45	618	9.1	1196	9	Q9ZX77	Q9zx77 clostridium

ALIGNMENTS

RESULT 1
Q9ZAJ8
ID Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

```
  Query Match             58.8%;  Score 4007;  DB 2;  Length 1291;
  Best Local Similarity   58.0%;  Pred. No. 1.6e-166;
  Matches  756;  Conservative  249;  Mismatches  280;  Indels   18;  Gaps   11;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | ||||||||:|::|||||     | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
               |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::|:||::| |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::|  |:|:|::|| |  ||  :: ||:  |||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
                  ||||  | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||:::||| || |:|||| :| :|||||||||| |||:|||||| :||:| ||||: |
Db        241 GIKVNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||| ||||:| : ||||||:|||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| :|||: |||:| : |||| |||  |||||: ||:: |: |||||:||:|
Db        361 IAENYKIKTRAAYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  |  |    || |:||||||||:|:||| ||:| | |||
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIAY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              ||||| |||:|||::|||| || | |:||:|||| |:| :: :||| || |:|||| |
Db        476 NTQNNYIENDFSINELILDTDLISKIELPSENTESLTDF-NVYVPVYKKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | ::|||  |: : ||:|||::|:|:|:|||| |||||||||| ||:||||| ||:||:||
Db        595 KQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | ||||: ||  || ||::|| |||:|| |||||||:|||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIETINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::: |||||||:: |||: || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYKYNIYSEKERSNINIDFNDVNSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              ||:||:|||| :||||||:|| |||: |:::|| ||| |:|||:      ||||:::||
Db        775 INECSVSYLMKKMIPLAVEKLLDFDNTLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
              ||||||| || :||||::|| | |:| :| ||:|  ||  :||| ||||| : |    |  |
Db        835 SIPFDLSTYTNNTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
              |    |||| :| || |   |:: ::::||| :||::||:| ||| |: || |: |||||
Db        895 D--KNQFKLTSSANSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
              |:|:||:||||:||:|| |||||||:| | ||:|||||||::||:|||:|| :||||:
Db        953 INCMKNNSGWKISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS- 1011

Qy       1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
              || ||||| |:    | ::  : ::::| |||    | |:|:|:| |:||  ||: : :
Db       1012 DNAKIYINGKLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
                :| ||| :  |||||||| |: :||:|| | :| |||    :|:||   |: :|  :
Db       1072 EEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEI 1197
                |||::||:| :|||:: |||::| ||:|||: |||||:  |:: | :|||:    |:
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYMYKYFKKE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
              : :||||||:|   ||: :||||:| |: ||:||:| :||   |   || || :|: :  |
Db       1190 EEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |:  | :|||||:|||: :         :|||||||| ||||||
Db       1249 VFKEYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 2
Q933K0
ID   Q933K0      PRELIMINARY;      PRT;  1291 AA.
AC   Q933K0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type B cryptic neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=593, and 588;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF300466; AAL11499.1; -.
DR   EMBL; AF300465; AAL11498.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1291 AA;  150843 MW;  7AC1737B0FA5A151 CRC64;

  Query Match             58.8%;  Score 4006;  DB 2;  Length 1291;
  Best Local Similarity   58.0%;  Pred. No. 1.7e-166;
  Matches  756;  Conservative  249;  Mismatches  280;  Indels   18;  Gaps   11;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | ||||||||:|::|||||     | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGMGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
               |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::|:||::| |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::|  |:|:|::|| |  ||  :: ||:  |||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180
```

```
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||  | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||:::||| || |:|||| :| :|||||||||| |||:|||||| :||:| |||| |
Db        241 GIKVNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||  ||||:| : ||||||:||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| |||| :||:| : |||| |||  ||||| ||:: |: |||||:||:|
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479
              ||||| ||| :|:| ||| |  |  |    || |:||||||||:|:||| ||:| | |||
Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIAY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539
              ||||| |||:|||::|||| || | |:||:|||| |:| :: :||| || |:|||| |
Db        476 NTQNNYIENDFSINELILDTDLISKIELPSENTESLTDF-NVYVPVYKKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599
              :::|:||::|||| :| :: ||:| :|||  :||||:||| : :: || || | ||  ||
Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659
              | :::||  |: : ||:||::|:|:|:|||| |||||||||| ||:|||| ||:||:||
Db        595 KQIVNDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719
              ||||::|:|| | ||||: ||  || ||::|| |||:|| |||||||:|||||||||||
Db        655 IPELLIPVVGAFLLESYIDNKNKIIETINSALTKRDEKWIDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779
              ||| || ||| |:||:|:||: :|| |||::: ||||||||:: |||: || ||:||::|
Db        715 IKEGMYKALNYQAQALEEIIKYKYNIYSEKERSNINIDFNDVNSKLNEGINQAIDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839
              ||:||:|||| :||||||:|| |||: |:::|| ||| |:|||:      ||||:::||
Db        775 INECSVSYLMKKMIPLAVEKLLDFDNTLRKNLLNYIDENKLYLIGSAEYEKSKVDKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
              ||||||| || :||||::|| | |:| :| ||:| ||  :||| ||||| : |   | |
Db        835 SIPFDLSTYTNNTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
              |    |||| :| || |   |:: ::::||| :||::||:| ||| |: || |: |||||
Db        895 D--KNQFKLTSSANSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
              |:|:||:||||:||:|| |||||||:| | ||:|||||||::||:|||:|| :||||:
Db        953 INCMKNNSGWKISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS- 1011

Qy       1020 GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
               || ||||| |:    | ::  : ::::| |||    | |:|:|:| |:||  ||: : :
Db       1012 DNAKIYINGKLESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
                :| ||| :  |||||||| |: :||:|| | :| |||    :|:||   |: :|  :
Db       1072 EEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
                |||::||:| :|||:: |||::| ||:||| |||||:  |:: | :|||: :|:.
Db       1132 KYINYRDLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYMYKYFKKE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
              : :||||||:|   ||: :|||:| |: ||:||:| :||   |   || || :| :  |
Db       1190 EEKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              |:  | :|||||:|||: :        :||||||||| ||||||
Db       1249 VFKEYKDYFCISKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 3
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match           58.7%;  Score 4001;  DB 2;  Length 1291;
  Best Local Similarity   58.0%;  Pred. No. 2.9e-166;
  Matches  756;  Conservative  235;  Mismatches  294;  Indels   18;  Gaps   10;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||| | |||||||:|::|||||    | | |||||:| |||||:|||:|:|::|: ||
Db          1 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
               |:|:|::|| ||||| || |: :|: ||:|||||||| ||| |::||:||:: |||||
Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180
              :   | ::|  |:|:|::|| |  ||  :: ||:  ||||||||||::| |  : :  |
Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240
              ||||  | ::|||  ::|||||||||  |||:|| ||:|||| ||||||||||||
Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300
              |||: :||| || |:|||| :| :|||||||||| |||:|:|||| :||:| |||: |
Db        241 GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359
              :||| ::       |:|::||  :|:|| |||  ||||:| : ||||||:||||||||
Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419
              :|  | |||| |||| :||:| : |||| |||  ||||| ||:: |: |||||:||:|
```

```
Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479

Db        421 YEEISKEHLAVYKIQMCKSVRAPGI-----CIDVDNEDLFFIADKNSFSDDLSKNERIEY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539

Db        476 DTQSNYIENRSSIDELILDTNLISKIELPSENTESLTDF-NVDVPVYEKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599

Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659

Db        595 KQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654

Qy        660 IPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYT 719

Db        655 IPELLIPVVGAFLLESYIDNKNKIIKTIDNALTKRDEKWIDMYGLIVAQWLSTVNTQFYT 714

Qy        720 IKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDF 779

Db        715 IKEGMYKALNYQAQALEEIIKYKYNIYSEKEKSNINIDFNDINSKLNEGINQAVDNINNF 774

Qy        780 INQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKD 839

Db        775 INECSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVDKHLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899

Db        835 IIPFDLSMYTNNTILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959

Db        895 D--KNQFKLTSSTNSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019

Db        953 INCIKNNSGWKISIRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNS- 1011

Qy       1020 GNANIYIINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079

Db       1012 DNAKIYIINGKLESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139

Db       1072 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197

Db       1132 NYINYRNLYIGEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFN-SNREWRVYAYKDFKEE 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254

Db       1190 EKKLFLANIYDSNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GI 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1249 VLKDYKNYFCISKWYLKEVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 4
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match           58.6%;  Score 3996;  DB 2;  Length 1291;
  Best Local Similarity 57.9%;  Pred. No. 4.7e-166;
  Matches  754;  Conservative  248;  Mismatches  283;  Indels   18;  Gaps   11;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db          1 MSVTINNFNYNDPIDNDNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN 60

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db         61 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH 180

Db        121 DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH 180

Qy        181 SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY 240

Db        181 FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY 240

Qy        241 GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA 300

Db        241 GIKVNDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIISPSTDKSIYDKVLQNFRGIV 300

Qy        301 NRLN-IVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETN 359

Db        301 DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN 360

Qy        360 LAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEA 419

Db        361 IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKNMEKEYRGQNKAINKQA 420

Qy        420 YEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAY 479

Db        421 YEEISKEHLAVYKIQMCKSV--KAPG---ICIDVDNEDLFFIADKNSFSDDLSKNERIAY 475

Qy        480 NTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDG 539

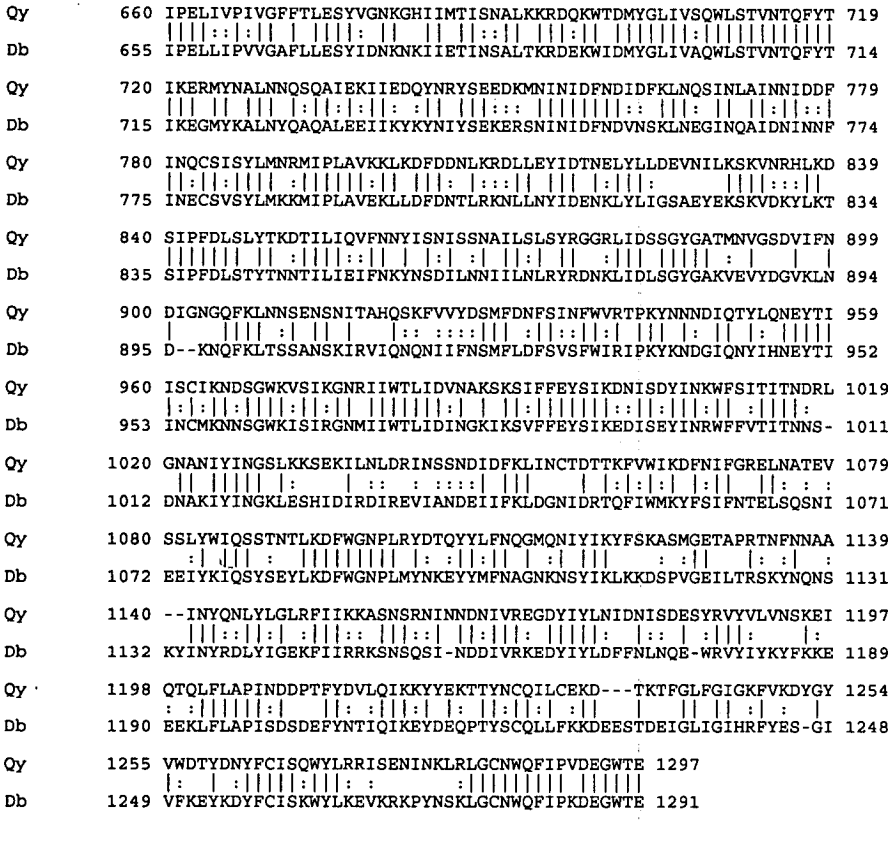
Db        476 NTQNNYIDNDFSINELILDTDLISKIELPSENTESLTDF-NVYVPEYKKQPAIKKIFTDE 534

Qy        540 DSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWV 599

Db        535 NTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWV 594

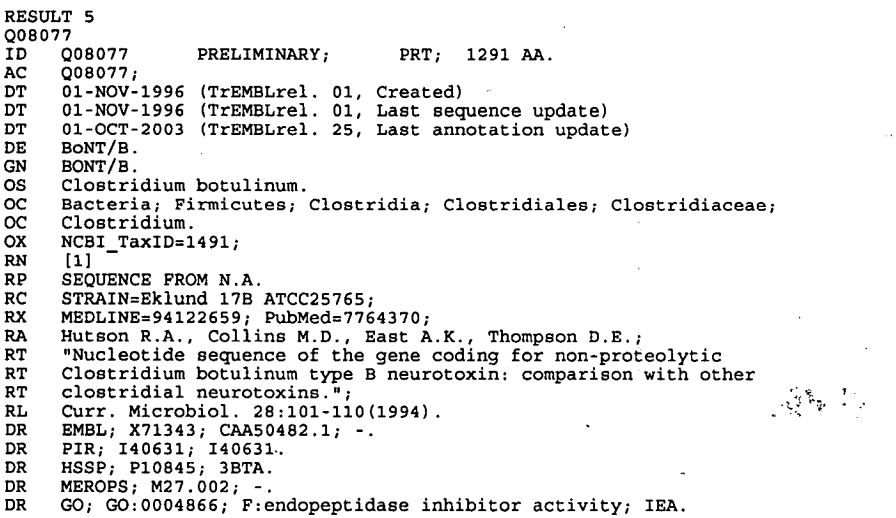
Qy        600 KGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEF 659

Db        595 KQIVDDFVIEANKSSTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEF 654
```



```
RESULT 5
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match            57.7%;  Score 3932;  DB 2;  Length 1291;
  Best Local Similarity  56.9%;  Pred. No. 2.9e-163;
  Matches  741;  Conservative  247;  Mismatches  297;  Indels   18;  Gaps   10;
```



```
Db        775 INECSVSYLMKKMIPLAVKKLLDFDNTLKKNLLNYIDENKLYLIGSVEDEKSKVDKYLKT 834

Qy        840 SIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFN 899
              |||||| |:   |||::|| | | | :| ||:| ||   ||| ||||| : |   |  |
Db        835 IIPFDLSTYSNIEILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLN 894

Qy        900 DIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTI 959
              |    |||| :| :| |   |:: :::::||| :||::||:| ||| |:||| |: |||||
Db        895 D--KNQFKLTSSADSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTI 952

Qy        960 ISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL 1019
              |:|:||:||||:||:|||||||||:| |:||:||||:|:::||:|||:|| :||||: |
Db        953 INCMKNNSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-L 1011

Qy       1020 GNANIYIINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEV 1079
              || |||||:|: :  | ::  :  : :| |||    | |:|:|:| |:||  :|| : :
Db       1012 DNAKIYINGTLESNMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNI 1071

Qy       1080 SSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA 1139
                :| ||| :  |||||||| |: :||:|| | :| |||    :|:||   |: :|  :
Db       1072 KEIYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNS 1131

Qy       1140 --INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 1197
                |||:|||:| :|||:: |||::| ||:|||: |||:|:: :  | :|||     ||
Db       1132 NYINYRNLYIGEKFIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQ 1189

Qy       1198 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGY 1254
              : :|.||: |:|   ||  ::||:| |: :|:||:| :||   |   || || :| :  |
Db       1190 EEKLFLSIISDSNEFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GV 1248

Qy       1255 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
              :   | :|||||:|||: :     |  |||||||| ||||||
Db       1249 LRKKYKDYFCISKWYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 6
Q93N27
ID   Q93N27      PRELIMINARY;      PRT;  1310 AA.
AC   Q93N27;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tetanus toxin (Fragment).
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shumin Z., Dianliang L.;
RT   "Cloning and sequence analysis of tetanus toxin gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF389424; AAK72964.2; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER       1      1
FT   NON_TER    1310   1310
SQ   SEQUENCE   1310 AA;  150316 MW;  9EADDC914418E450 CRC64;

  Query Match           36.5%;  Score 2488.5;  DB 2;  Length 1310;
  Best Local Similarity 39.6%;  Pred. No. 2.2e-100;
  Matches  532;  Conservative  253;  Mismatches  473;  Indels   85;  Gaps   25;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              :|: | || |: |:||| |||||     |   |||||:| |||||||||: :| :|: ||
Db          2 IPITINNFRYSVPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN 61

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                : :  :  ||||| ||:||::||:||:||:|||||| :  :|: ||| |::||||||
Db         62 PPSSLI-EGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLG 120

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIM--- 177
              |: :  |||  |  :|| |      |         ::|:||||||||||: |    :::
Db        121 NSYSLLDKFDTNSNSVSFNLSEQDPSGATTKSAMLTSLIIFGPGPVLNKNEVRGIVLRVD 180

Qy        178 -NGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVL 236
               : |  :|||: | : |||  :  |:|| ||  :    :  || |||| |||||||||
Db        181 NKNYFPCRDGFGSIMQMTFCPEYIPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVL 240

Qy        237 HGLYGIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNF 296
              |||||:::|:  | |: :| :|||: |: ||||:|||| | ::||     ::| | | ::
Db        241 HGLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDY 300

Qy        297 QDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFT 356
              : |||:|: |:|     |||  ||||| || | :|  ||:|  |::|||  || ::|:|||
Db        301 KAIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFT 360

Qy        357 ETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVN 416
              |  |  :: |||| ||||    |:|   |||:|||   ||||| ||:||:|: |||  ||
Db        361 EIELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVN 420

Qy        417 KEAYEEISLEHLVIYRIAMCKPV---------MYKNTGK-----SEQCIIVNNEDLFFIA 462
              |:  :    ||   | :|| :          :|  |          | || : |||| |||
Db        421 TNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTAALTDLGGELCIKIKNEDLIFIA 480

Qy        463 NKDSFSKDLAKAETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDID 522
              |:|||::  : | ::|||:|  :  |:|:|::||| :| | | ||:| | |    |
Db        481 EKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIILDYNLQSKITLPNDRTTPVTK--GIP 538

Qy        523 IPVYIKQSALKKI---FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFS 579
                  | :|    |    :| :::::||:||  |: :: : :|||::||| |: |:|::|
Db        539 YAPEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYF- 597

Qy        580 TNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNET 639
               ::: | |       ||: ||: :|||||:||:||:||||:|||| |:||||||||:  :
Db        598 PSVISKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQG 657

Qy        640 AKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWT 699
              : ||  | |  |  :|:|:|||: :|::   ::       |  || || | |:|| :||
Db        658 YEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWI 717

Qy        700 DMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFN 759
              ::| |: ::|| ||||||      :|| :|  |  ||:|||: :|  ||  ||  |  : |
Db        718 EVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEIN 777

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819
              ::  || :  | |: ||: |: : | |:|:|:||    |:| :||   |  |::||  |
Db        778 NLKNKLEEKANKAMININIFMRESSRSFLVNQMINETKKQLLEFDTQSKNILMQYIKANS 837

Qy        820 LYL-LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISN-------ISSNAIL 871
              :: : |:  |:||:|:      |||     |:|:       : :: |       :  : ||
Db        838 KFIGITELKKLESKINKVFSTPIPFS---YSKN------LDCWVDNEEDIDVILKKSTIL 888

Qy        872 SLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMF 930
              :|      : | ||: :::     |  :    |       | |:|:|  :  |::  :  |: ||
Db        889 NLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMF 948

Qy        931 DNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRIIWTLIDV 985
              :|:::||:| || : : :: |  |||:||| :|       ||| ||:||| :|||| |
Db        949 NNFTVSFWLRVPKVSASHLEQYDTNEYSIISSMKKYSLSIGSGWSVSLKGNNLIWTLKDS 1008

Qy        986 NAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINS 1044
```

```
                 : : | |   : |  : |: |||  |||||||| :||:|||| |  | :|  |  |
Db       1009 AGEVRQITFR-DLSDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIRE 1067

Qy       1045 SNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDT 1104
                |:|  ||  | :  ::| |  | || : ||  |:  ||    |   |:||||||||||
Db       1068 DNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDT 1127

Qy       1105 QYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFIIKKASNSR 1160
                :|||       ::: :|  :       ||  :: |  :|  |: || ||:||||: : :
Db       1128 EYYLIPVAYSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYSGLKFIIKRYTPNN 1185

Qy       1161 NINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPINDDPTFYDV 1215
                 |  |: || ||:| | :   ::|    |        |    |    : ||      |
Db       1186 EI--DSFVRSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP---GIPLYKK 1240

Qy       1216 LQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNYFCISQWY 1269
                ::  |  :  ||: |: | :    : || |     ||          |   :    | ||
Db       1241 MEAVKLRDLKTYSVQLKLYDDKDASLGLVGTHNGQIGN---------DPNRDILIASNWY 1291

Qy       1270 LRRISENINKLRLGCNWQFIPVD 1292
                 : :      | |:| |:| |
Db       1292 FNHLKDK----TLTCDWYFVPTD 1310


RESULT 7
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match             34.2%;  Score 2328.5;  DB 2;  Length 1268;
  Best Local Similarity   38.6%;  Pred. No. 1.9e-93;
  Matches  519;  Conservative  252;  Mismatches  450;  Indels  123;  Gaps   34;

Qy          1 MPVNIKNFNYNDPINNDDIIMME-PFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQF 59
              ||||| ||||||||||  |: |: |: :     ||||| |:| :|:|||   | :|  |
Db          1 MPVNINNFNYNDPINNTTILYMKMPYYEDS-NKYYKAFEIMDNVWIIPERNIIGKKPSDF 59

Qy         60 NASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
                 : |     |||| || ||||||:||||:||||||||| |:|| ||: | :  |||
Db         60 YPPISLDSGS-SAYYDPNYLTTDAEKDRFLKTVIKLFNRINSNPAGQVLLEEIKNGKPYL 118

Qy        120 GNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD--NFTDSMIM 177
              ||  |  ::| ||  : |:  |    |  |    :: ||:| |||| : :   |   :
Db        119 GNDHTAVNEFCANNRSTSVEIK-ESNGTTD---SMLLNLVILGPGPNILECSTFPVRIFP 174

Qy        178 N--GHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHV 235
              |   : |  :|||:  :: |      ||   :| |         : ||||::| ||||||
Db        175 NNIAYDPSEKGFGSIQLMSFSTEYEYAFN---DNTD-------LFIADPAISLAHELIHV 224

Qy        236 LHGLYGIK-ISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQ 294
              |||||| | ::|  :    :    |      :: ||  |||| | ::|: ||:  ||   |
Db        225 LHGLYGAKGVTNKKVIEVDQGALMAAEKDIKIEEFITFGGQDLNIITNSTNQKIYVILLS 284

Qy        295 NFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFG 354
              |:  ||:||: | :   | :: : ||  :: ||   :| || |:|:  ||: :|| | |
Db        285 NYTAIASRLSQV-NRNNSALNTTYYKNFFQWKYGLDQDSNGNYTVNISKFNAIYKKL-FS 342

Qy        355 FTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKA 414
              ||| :|| :: :| | :|   :  | :   |||: ||: :||||| |  |:   ||||
Db        343 FTECDLAQKFQVKNRSNYLFHF-KPFRLLDLLDDNIYSISEGFNIGS--LRVNNNGQNIN 399

Qy        415 VNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKA 474
              :|      |    ||   : :|| :|  |:  || ||| ||||:|:: |::::   :
Db        400 LNSRIVGPIPDNGLVERFVGLCKSIVSKKGTKNSLCIKVNNRDLFFVASESSYNENGINS 459

Qy        475 ETIAYNTQNNTIENN---FSIDQLILDNDLSSGIDLPNENTEPF--TNFDDIDIPVYIKQ 529
                    : || ||    ::|::||| :  :   :|| ::     |  :|  :| |
Db        460 PK---EIDDTTITNNNYKKNLDEVILDYNSDA---IPNLSSRLLNTTAQNDSYVPKYDSN 513

Qy        530 --SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKAN 587
                | :|:  ||  ::| ||:||  |     : ||:|:| || : :|||||||:: :   |
Db        514 GTSEIKEYTVDKLNVFFYLYAQKAPEGESAISLTSSVNTALLDASKVYTFFSSDFINTVN 573

Qy        588 TVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNA 647
                | :||::|:: ||:|||:|:|||||||::|:|:|:||:| |||:|||  | ||| |
Db        574 KPVQAALFISWIQQVINDFTTEATQKSTIDKIADISLIVPYVGLALNIGNEVQKGNFKEA 633

Qy        648 FEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGL 704
                |: || ||:||:|||::|  :  ||::|::    :|  ||  |:|||::|: || ::|
Db        634 IELLGAGILLEFVPELLIPTILVFTIKSFINSDDSKNKIIKAINNALRERELKWKEVYSW 693

Qy        705 IVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFN--DID 762
              ||| ||: :||||   ||:|| || ||   |:|||| :|| |: ::|  :  ::|   |
Db        694 IVSNWLTRINTQFNKRKEQMYQALQNQVDGIKKIIEYKYNNYTLDEKNRLRAEYNIYSIK 753

Qy        763 FKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYL 822
              :||: ::||: ||| |: : ||||||  :    : || ::|  : : || ||  |   |
Db        754 EELNKKVSLAMQNIDRFLTESSISYLMKLINEAKINKLSEYDKRVNQYLLNYILENSSTL 813

Qy        823 -LDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLI 881
                 |  | : |:  | :||||:|| || | ||| :   :    |  ::||:: |   | |
Db        814 GTSSVPELNNLVSNTLNNSIPFELSEYTNDKILIHILIRFYKRIIDSSILNMKYENNRFI 873

Qy        882 DSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRT 941
              ||||||: :::  |:         || : :|  | :   |:  ::|:| : |||::||||
Db        874 DSSGYGSNISINGDIYIYSTNRNQFGIYSSRLSEVNITQNNTIIYNSRYQNFSVSFWVRI 933

Qy        942 PKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKD 1000
              |||||       | |||||:|:: |:||||:|:  | |||| |    :: : | |:
Db        934 PKYNN---LKNLNNEYTIINCMRNNNSGWKISLNYNNIIWTLQDTTGNNQKLVFNYTQMI 990

Qy       1001 NISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTK 1060
              :|||||||  :||||:|||:: :||||:|   : ||||  |:  ::| ||:: | | |:
Db        991 DISDYINKWTFVTITNNRLGHSKLYINGNLTDQKSILNLGNIHVDDNILFKIVGCND-TR 1049

Qy       1061 FVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK- 1119
              :| |: | ||  ||: ||: :||  :  :  |||||| | |: :||| |    |: :
Db       1050 YVGIRYFKIFNMELDKTEIETLYHSEPDSTILKDFWGNYLLYNKKYYLLNLLKPNMSVTK 1109

Qy       1120 ------------YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNI 1167
```

```
                        :|| ::        |:||      || |:  ||:|  ::   | ||
Db       1110 NSDILNINRQRGIYSKTNI--------FSNA-----RLYTGVEVIIRKVGSTDTSNTDNF 1156

Qy       1168 VREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTY 1227
              ||: | :|:|: : : | |::|  |::  ::  : |  |::                : |
Db       1157 VRKNDTVYINVVDGNSE-YQLYADVSTSAVEKTIKLRRISN---------------SNY 1199

Qy       1228 NCQIL---------CEKDTKT-----FGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRI 1273
              |   :          |  : ||      || |           : |    | || : |
Db       1200 NSNQMIIMDSIGDNCTMNFKTNNGNDIGLLGF------------HLNNLVASSWYYKNI 1246

Qy       1274 SENINKLRLGCNWQFIPVDEGWTE 1297
                |      || | ||  : || |
Db       1247 RNNTR--NNGCFWSFISKEHGWQE 1268
```

```
RESULT 8
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           33.9%;  Score 2308;  DB 2;  Length 1278;
  Best Local Similarity  38.2%;  Pred. No. 1.5e-92;
  Matches  510;  Conservative  255;  Mismatches  475;  Indels   94;  Gaps   32;

Qy          1 MPVNIKNFNYNDPINNDDIIMME-PFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQF 59
              ||| | :|||||:|:| |: |: |: :      |||| |:  :||:||| | |  | |
Db          1 MPVVINSFNYNDPVNDDTILYMQIPYEEKSK-KYYKAFEIMRNVWIIPERNTIGTDPSDF 59

Qy         60 NASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYL 119
               :    :      |||| || ||||||::||| |||| |||| |:|: ||  |  | |||
Db         60 DPPASL-ENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYL 118

Qy        120 GNASTPPDKFAANVANVSINKKIIQPGAEDQIK-GLMTNLIIFGPGPVLSDN--FTDSMI 176
              ||  || ::|       |:| |     :   :|  :: ||:: | || : :|  :    :
Db        119 GNEHTPINEFHPVTRTTSVNIK-----SSTNVKSSIILNLLVLGAGPDIFENSSYPVRKL 173

Qy        177 MNG---HSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELI 233
              |:    : | ::|||:  :: | |      ||::    ::   |  :: ||||::| ||||
Db        174 MDSGGVYDPSNDGFGSINIVTFSPEYEYTFNDISGGYNS---STESFIADPAISLAHELI 230

Qy        234 HVLHGLYG---------IKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPST 284
              | ||||||         ||:   |:        |    |:: ||  |||| | ::|: :
Db        231 HALHGLYGARGVTYKETIKVKQAPL--------MIAEKPIRLEEFLTFGGQDLNIITSAM 282

Qy        285 DMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKF 344
              |||  | |:: || ||: |:||      ||: ||  :: ||   :: :| |:|:::||
Db        283 KEKIYNNLLANYEKIATRLSRVNSAPPE-YDINEYKDYFQWKYGLDKNADGSYTVNENKF 341

Qy        345 DKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNL 404
              :::|| | : ||| :|| :: :| | :|| :|   :|   |||: ||| :|||||   ||
Db        342 NEIYKKL-YSFTEIDLANKFKVKCRNTYFIKY-GFLKVPNLLDDDIYTVSEGFNIG--NL 397

Qy        405 KTEFNGQNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKS--EQCIIVNNEDLFFIA 462
                  |||  :| :  : |  : ||   :  || |: :   |:    || ||| :|||:|
Db        398 AVNNRGQNIKLNPKIIDSIPDKGLVEKIVKFCKSVIPRKGTKAPPRLCIRVNNRELFFVA 457

Qy        463 NKDSFSK-DLAKAETIAYNTQ-NNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDD 520
              :: |::: |:   : |   | ||  ||  :|::||| :  :   : |:          ||
Db        458 SESSYNENDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSETIPQISNQTLNTLVQ-DD 514

Qy        521 IDIPVYIKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFF 578
              :| |    | :::  |   ::| |||||  |    |: ||:|:: ||   ::|||||
Db        515 SYVPRYDSNGTSEIEEHNVVDLNVFFYLHAQKVPEGETNISLTSSIDTALSEESQVYTFF 574

Qy        579 STNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNE 638
              |:  :   |  | |:||::|:  || |||:|:|||||  ||::|:|:::||:| |||:|||
Db        575 SSEFINTINKPVHAALFISWINQVIRDFTTEATQKSTFDKIADISLVVPYVGLALNIGNE 634

Qy        639 TAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRD 695
              |||||  |||: || ||:||:|||::|  :  ||::|::|   ||  ||  |:|:| :|:
Db        635 VQKENFKEAFELLGAGILLEFVPELLIPTILVFTIKSFIGSSENKNKIIKAINNSLMERE 694

Qy        696 QKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNIN 755
              || ::|  ||| ||: :||||   ||:|| || ||  ||: :|| :|| |: ::: :
Db        695 TKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTVIEYKYNNYTSDERNRLE 754

Qy        756 IDF--NDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLE 813
                ::  |:|  :||: ::||: ||: || : || |||  :    | ||:::|: :|  ||:
Db        755 SEYNINNIREELNKKVSLAMENIERFITESSIFYLMKLINEAKVSKLREYDEGVKEYLLD 814

Qy        814 YIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSL 873
              ||  :  | : |  |    |  :|||:|| || | ||| ||    |  |:|| :
Db        815 YISEHRSILGNSVQELNDLVTSTLNNSIPFELSSYTNDKILILYFNKLYKKIKDNSILDM 874

Qy        874 SYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNF 933
              |   : || ||||: :::  ||        || : :|: | :   |:  ::|: : ||
Db        875 RYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQNF 934
```

```
Qy        934 SINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSI 992
              ||:|||| ||| |   :  | |||||| ||: |:||||:|:  |:||||| |    :: :
Db        935 SISFWVRIPKYFN---KVNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQDTAGNNQKL 991

Qy        993 FFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKL 1052
              | |:   :|||||||  :||||:||||: |||||:|   : | ||  |: |::| ||:
Db        992 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 1051

Qy       1053 INCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQG 1112
              : | | |::| |: | :|  ||  ||: :||  :   : |||||| | |: :||| |
Db       1052 VGCND-TRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLN-- 1108

Qy       1113 MQNIYIKYFSKASMGETAPRTNFNNAAINYQ--------NLYLGLRFIIKKASNSRNINN 1164
                    :  |: : :   | |     ||         || |:  ||:|  ::   |
Db       1109 ------LLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVIIRKNGSTDISNT 1162

Qy       1165 DNIVREGDYIYLNIDNISDESYRVYVLVN-SKEIQTQLFLAPINDDPTFYDVLQIKKYYE 1223
              || ||: | |:|: :  |  ||:|  :: :|  :    :   | : :   :: :
Db       1163 DNFVRKNDLAYINVVD-RDVEYRLYADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGN 1221

Qy       1224 KTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLG 1283
               | | |     :   || |           : |    | ||   | :| :    |
Db       1222 NCTMNFQ---NNNGGNIGLLGF----------HSNNLVASSWYYNNIRKNTSS--NG 1263

Qy       1284 CNWQFIPVDEGWTE 1297
              | | ||  : || |
Db       1264 CFWSFISKEHGWQE 1277


RESULT 9
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
```

```
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             33.1%;  Score 2259.5;  DB 2;  Length 1251;
  Best Local Similarity   38.4%;  Pred. No. 1.9e-90;
  Matches  506;  Conservative  262;  Mismatches  460;  Indels   91;  Gaps   33;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||  | :|||||:|:  |: ::|    |   :||:| |:  |||:|||    |  |  |:
Db          1 MP-KINSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFH 56

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                | : : |   |||| ||::| |||:||| : |:|||||:  ||  ||: :  | ||||
Db         57 PPTSLKNGD-SSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG 115

Qy        121 NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMNG 179
              | :|| ::|  :: : |  :     |::|    |: |:|| |  | | : | ::  : |
Db        116 NDNTPDNQF--HIGDASAVEIKFSNGSQDI---LLPNVIIMGAEPDLFETNSSNISLRNN 170

Qy        180 HSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGL 239
              : | : |||:  :: | |    ||:   |:          :  ||||||||||| ||||
Db        171 YMPSNHGFGSIAIVTFSPEYSFRFNDNSMNE---------FIQDPALTLMHELIHSLHGL 221

Qy        240 YGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNKA 292
              || |     ||  ||    |  :|:          ||  |||| | ::|: :   :||
Db        222 YGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNL 275

Qy        293 LQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALM 352
              | ::: ||::|   |  | |   :: || ::: ||   :| :| |||: :||: ::| |
Db        276 LADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL- 331

Qy        353 FGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQN 412
              : ||| :|| :: :| | :|  :|   |  ||:::||  :||:||   |||  | |||
Db        332 YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQN 388

Qy        413 KAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLA 472
              :|      |:   ||   |  || ::    :   || :|| :|||:|:::|:: |
Db        389 ANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNI 448

Qy        473 KA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYI 527
                    :|:   | ||  ||:  :||:||: :  |   | :|     |  :|  || |
Db        449 NTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKYD 502

Qy        528 KQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEK 585
              |  :::  |:  ::| || ||   |     |: ||:|:: ||    |:|||:  :
Db        503 SNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINN 562

Qy        586 ANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFK 645
              |  | |:|||:|:: |: |||:|: ||||:||::|:|::|||| |||:|||  | |||
Db        563 VNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFK 622

Qy        646 NAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMY 702
              :| |: || ||:||:|||::| :  ||::|::|   ||  :|  |:||||:||:|| ::|
Db        623 DALELLGAGILLEFVPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVY 682

Qy        703 GLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFND 760
              ||| |:: :|||   ||:|| || ||  |:: ||| :|| |: |:|  :  | |
Db        683 SFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKTIIEFKYNSYTLEEKKELKNNYDIEQ 742

Qy        761 IDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNEL 820
              |: :||| :::|:|||| |: : |||||  :  : : ||:::|:|:|  ||:||  :
Db        743 IENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIQHGS 802

Qy        821 YLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRL 880
              | :    | | |   | :|||| || || | |||  || :   | |:::|:: |:  :
Db        803 ILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKY 862

Qy        881 IDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVR 940
              :|:||| : :|:  ::          || : |:: | :   |: :::||: : ||||:|||
Db        863 VDTSGYDSNININGEIFIYPTNKNQFTIFNSKPSEVNISQNDYIIYDNKYKNFSISFWVR 922

Qy        941 TPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIK 999
              | |:|  :   : |||||:|:: |:||||||:  | |||||  :  :: : |:|
Db        923 IPNYDNKIVN--INNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNARINQKLVFKYGNA 980

Qy       1000 DNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTT 1059
              : |||||||  :||||||||:: :|||| |   : ||||  |: |::| ||::||:  |
Db        981 NGISDYINKWIFVTITNDRLGDSKLYINGHLIDQKSILNLGNIHVSDNILFKIVNCS-YT 1039

Qy       1060 KFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK 1119
              ::: |: |||| :||: ||: :||  : :|| ||||||| | ||  ||| |    | :|
Db       1040 RYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKGYYLLNVLKPNNFI- 1098

Qy       1120 YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLN-I 1178
                        |    |  :  :    || |::  |:: ::|    ||  ||: | :|:| |
Db       1099 ---DRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNDSS--TNDRFVRKNDQVYINYI 1153

Qy       1179 DNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTK 1238
                | |  || :|   |: :   :  :  :    |  |: :          | | :      :
Db       1154 SNSS--SYSLYADTNTTD--KEKTIKSSSSGNRFNQVVVMNSVGNNCTMNFK---NNNGN 1206

Qy       1239 TFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297
                ||          |: ||       | ||   : :: |    || | ||  : || |
Db       1207 NIGLL---------GFKADT----VVASTWYYTHMRDHTNS--NGCFWNFISEEHGWQE 1250


RESULT 10
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match             32.8%;  Score 2234;  DB 2;  Length 1252;
  Best Local Similarity   38.3%;  Pred. No. 2.5e-89;
  Matches  505;  Conservative  256;  Mismatches  468;  Indels   90;  Gaps   33;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60
              ||  | :||||||:||  |: ::|    |   :||:| |:  |||:|||    |  |  |
Db          1 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 56

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120
                | :. |   |||| ||::| ||||||| : |:|||||   ||: ||: :  | ||||
Db         57 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 115

Qy        121 NASTPPDKFAANVAN-VSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMN 178
```

```
Db        116 NDNTPDGDFIINDASAVPIQ---FSNGSQSI---LLPNVIIMGAEPDLFETNSSNISLRN 169

Qy        179 GHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHG 238

Db        170 NYMPSNHGFGSIAIVTFSPEYSFRF------KDNSM---NEFIQDPALTLMHELIHSLHG 220

Qy        239 LYGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNK 291

Db        221 LYGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTN 274

Qy        292 ALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKAL 351

Db        275 LLADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL 331

Qy        352 MFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQ 411

Db        332 -YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQ 387

Qy        412 NKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDL 471

Db        388 NANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDN 447

Qy        472 AKA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVY 526

Db        448 INTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKY 501

Qy        527 IKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVE 584

Db        502 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 561

Qy        585 KANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENF 644

Db        562 NVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 621

Qy        645 KNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDM 701

Db        622 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 681

Qy        702 YGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFN 759

Db        682 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIE 741

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819

Db        742 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHG 801

Qy        820 LYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGR 879

Db        802 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 861

Qy        880 LIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWV 939

Db        862 YVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWV 921

Qy        940 RTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSI 998

Db        922 RIPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGN 979

Qy        999 KDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDT 1058

Db        980 ANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCS-Y 1038

Qy       1059 TKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYI 1118

Db       1039 TRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFI 1098

Qy       1119 KYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI 1178

Db       1099 ----NRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYINF 1152

Qy       1179 DNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTK 1238

Db       1153 --VASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFK---NNNGN 1207

Qy       1239 TFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1208 NIGLL---------GFKADT----VVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1251
```

RESULT 11
Q9FAR6
ID Q9FAR6 PRELIMINARY; PRT; 1255 AA.
AC Q9FAR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX MEDLINE=20509829; PubMed=11055954;
RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT butyricum Strains.";
RL Appl. Environ. Microbiol. 66:4992-4997(2000).
DR EMBL; AB039264; BAB12249.1; -.
DR HSSP; P10845; 3BTA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CD8E4D CRC64;

Query Match 32.8%; Score 2234; DB 2; Length 1255;
Best Local Similarity 38.3%; Pred. No. 2.5e-89;
Matches 505; Conservative 256; Mismatches 468; Indels 90; Gaps 33;

```
Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN 60

Db          4 MP-TINSFNYNDPVNNRTILYIKP---GGCQQFYKSFNIMKNIWIIPERNVIGTIPQDFL 59

Qy         61 ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG 120

Db         60 PPTSLKNGD-SSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG 118

Qy        121 NASTPPDKFAANVAN-VSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSD-NFTDSMIMN 178

Db        119 NDNTPDGDFIINDASAVPIQ---FSNGSQSI---LLPNVIIMGAEPDLFETNSSNISLRN 172

Qy        179 GHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHG 238

Db        173 NYMPSNHGFGSIAIVTFSPEYSFRF------KDNSM---NEFIQDPALTLMHELIHSLHG 223

Qy        239 LYGIKISNLPITPNTKEFFMQHSDPV-------QAEELYTFGGHDPSVISPSTDMNIYNK 291

Db        224 LYGAK----GIT--TKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTN 277

Qy        292 ALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKAL 351
```

```
Db        278 LLADYKKIASKL---SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKL 334

Qy        352 MFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQ 411

Db        335 -YSFTEFDLATKFQVKCRQTYIGQY-KYFKLSNLLNDSIYNISEGYNI--NNLKVNFRGQ 390

Qy        412 NKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDL 471

Db        391 NANLNPRIITPITGRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDN 450

Qy        472 AKA-----ETIAYNTQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVY 526

Db        451 INTPKEIDDTV---TSNNNYEND--LDQVILNFNSESAPGLSDEKLN-LTIQNDAYIPKY 504

Qy        527 IKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVE 584

Db        505 DSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFIN 564

Qy        585 KANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENF 644

Db        565 NVNKPVQAALFVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNF 624

Qy        645 KNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDM 701

Db        625 KDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEV 684

Qy        702 YGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNI--NIDFN 759

Db        685 YSFIVSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIE 744

Qy        760 DIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNE 819

Db        745 QIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIKHG 804

Qy        820 LYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGR 879

Db        805 SILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDK 864

Qy        880 LIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWV 939

Db        865 YVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWV 924

Qy        940 RTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSI 998

Db        925 RIPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGINQKLAFNYGN 982

Qy        999 KDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDT 1058

Db        983 ANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSDNILFKIVNCS-Y 1041

Qy       1059 TKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYI 1118

Db       1042 TRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDKEYYLLNVLKPNNFI 1101

Qy       1119 KYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI 1178

Db       1102 ----NRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYINF 1155

Qy       1179 DNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTK 1238

Db       1156 --VASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFK---NNNGN 1210

Qy       1239 TFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 1297

Db       1211 NIGLL---------GFKADT----VVASTWYYTHMRDNTNS--NGFFWNFISEEHGWQE 1254


RESULT 12
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;

  Query Match             32.7%;  Score 2230.5;  DB 2;  Length 1280;
  Best Local Similarity   37.8%;  Pred. No. 3.6e-89;
  Matches  505;  Conservative  260;  Mismatches  477;  Indels   93;  Gaps   32;

Qy          1 MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQF- 59

Db          1 MPVVINSFNYNDPVNDETILYMQKPYEERSRKYYKAFEIMPNVWIMPERDTIGTKPDEFQ 60

Qy         60 ------NASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIV 113

Db         61 VPDSLKNGSSA--------YYDPNYLTTDAEKDRYLKTMIKLFNRINSNPTGKVLLEEVS 112

Qy        114 DAIPYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPG--------- 164

Db        113 NARPYLGDDDTLINEFLPVNVTTSVNIKF----STDVESSIISNLLVLGAGPDIFKAYCT 168

Qy        165 PVLSDNFTDSMIMNGHSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADP 224

Db        169 PLVRFNKSDKLI----EPSNHGFGSINILTFSPEYEHIFNDISGGNHN---STESFIADP 221

Qy        225 ALTLMHELIHVLHGLYGIK-ISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPS 283

Db        222 AISLAHELIHALHGLYGAKAVTHKESLVAERGPLMIAEKPIRLEEFLTFGGEDLNIIPSA 281

Qy        284 TDMNIYNKALQNFQDIANRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDK 343

Db        282 MKEKIYNDLLANYEKIATRLREVNTAP-PGYDINEYKDYFQWKYGLDRNADGSYTVNRNK 340

Qy        344 FDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKN 403

Db        341 FNEIYKKL-YSFTEIDLANKFKVKCRNTYFIKY-GFVKVPNLLDDDIYTVSEGFNIG--N 396

Qy        404 LKTEFNGQNKAVNKEAYEEISLEHLVIYRIAMCKPVM-YKNTGKSEQ-CIIVNNEDLFFI 461

Db        397 LAVNNRGQNINLNPKIIDSIPDKGLVEKIIKFCKSIIPRKGTKQSPSLCIRVNNRELFFV 456

Qy        462 ANKDSFSK-DLAKAETIAYNTQ-NNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFD 519
```

```
Db        457 ASESSYNESDINTPKEIDDTTNLNNNYRNN--LDEVILDYNSETIPQISNRTLNTLVQ-D 513

Qy        520 DIDIPVYIKQ--SALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTF 577

Db        514 NSYVPRYDSNGTSEIEEYDVVDFNVFFYLHAQKVPEGETNISLTSSIDTALLEESKVYTF 573

Qy        578 FSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGN 637

Db        574 FSSEFIDTINKPVNAALFIDWISKVIRDFTTEATQKSTVDKIADISLIVPYVGLALNIVI 633

Qy        638 ETAKENFKNAFEIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKR 694

Db        634 EAEKGNFEEAFELLGAGILLEFVPELTIPVILVFTIKSYIDSYENKNKAIKAINNSLIER 693

Qy        695 DQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYS--EEDKM 752

Db        694 EAKWKEIYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDAIKTAIEYKYNNYTSDEKNRL 753

Qy        753 NINIDFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLL 812

Db        754 ESKYNINNIEEELNKKVSLAMKNIERFMTESSISYLMKLINEAEVGKLKEYDKHVKSDLL 813

Qy        813 EYIDTNELYLLDEVNILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYISNISSNAILS 872

Db        814 DYILYHKLILGEQTKELIDLVTSTLNSSIPFELSSYTNDKILIIYFNRLYKKIKDSSILD 873

Qy        873 LSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDN 932

Db        874 MRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYSGRLSEVNIAQNNDIIYNSRYQN 933

Qy        933 FSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KNDSGWKVSIKGNR---IIWTLID 984

Db        934 FSISFWVTIPKHYRPMNRN-------REYTIINCMGNNNSGWKISLRTIRDCEIIWTLQD 986

Qy        985 VNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINS 1044

Db        987 TSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGNSRIYINGNLIVEKSISNLGDIHV 1046

Qy       1045 SNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDT 1104

Db       1047 SDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETLYSNEPDPSILKDYWGNYLLYNK 1105

Qy       1105 QYYLFNQGMQNIYIKYFSK-ASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNIN 1163

Db       1106 KYYLFNLLRKDKYITRNSGILNINQQRGVTGGISVFLNYK-LYEGVEVIIRKNAPIDISN 1164

Qy       1164 NDNIVREGDYIYLNIDNISDESYRVYVLVN-SKEIQTQLFLAPINDDPTFYDVLQIKKYY 1222

Db       1165 TDNFVRKNDLAYINVVDHGVE-YRLYADISITKSEKIIKLIRTSNPNDSLGQIIVMDSIG 1223

Qy       1223 EKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRL 1282

Db       1224 NNCTMNFQ---NNDGSNIGLLGF------------HSDDLVASSWYYNHIRRNTSS--N 1265

Qy       1283 GCNWQFIPVDEGWTE 1297

Db       1266 GCFWSFISKEHGWKE 1280


RESULT 13
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match             30.5%;  Score 2077;  DB 12;  Length 1275;
  Best Local Similarity   35.3%;  Pred. No. 1.7e-82;
  Matches  465;  Conservative  255;  Mismatches  486;  Indels  110;  Gaps   26;

Qy          5 IKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFNASTG 64

Db          5 VKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPS------- 57

Qy         65 VFSK------DVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPY 118

Db         58 -LSKPPRPTSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPF 116

Qy        119 LGNASTPPD--KFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMI 176

Db        117 MGDSSTPEDTFDFTRHTTNIAVEK--FENGSWKVTNIITPSVLIFGPLPNILD-YTASLT 173

Qy        177 MNGH--SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIH 234

Db        174 LQGQQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTH 233

Qy        235 VLHGLYGIKI-SNLPITPNTKE-FFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKA 292

Db        234 SLHQLYGINIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKA 293

Qy        293 LQNFQDIANRLNIVSSAQGSG--IDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKA 350

Db        294 LGHYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSD 353

Qy        351 LMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNG 410

Db        354 LTNVMSEVVYSSQYNVKNRTHYFSRHYLPV-FANILDDNIYTIRDGFNLTNKGFNIENSG 412

Qy        411 QNKAVNKEAYEEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKD 470

Db        413 QNIERN-PALQKLSSESVVDLFTKVCLR-LTKNSRDDSTCIKVKNNRLPYVADKDSISQE 470

Qy        471 LAKAETIAYNTQNNTIENNFSIDQLILDNDLSSG---ID--LPNENTEP-------FTNF 518

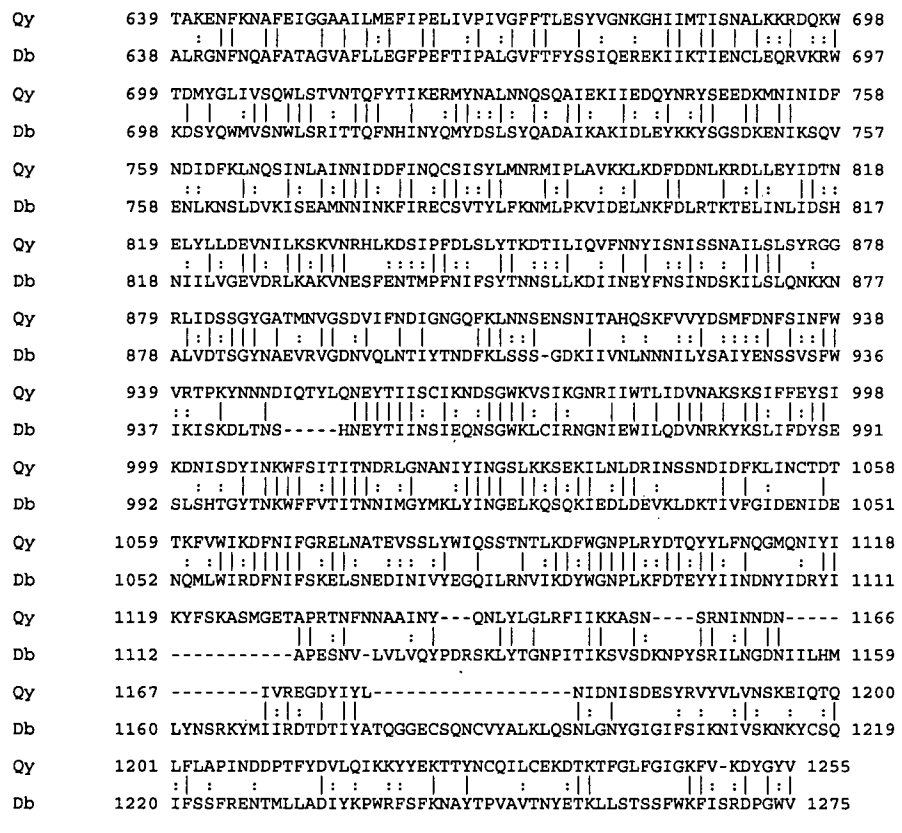
Db        471 IFENKIITDETNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFY 530

Qy        519 DDIDIPVYIKQSALKKIFVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFF 578

Db        531 DDI------------TKYVDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFL 578

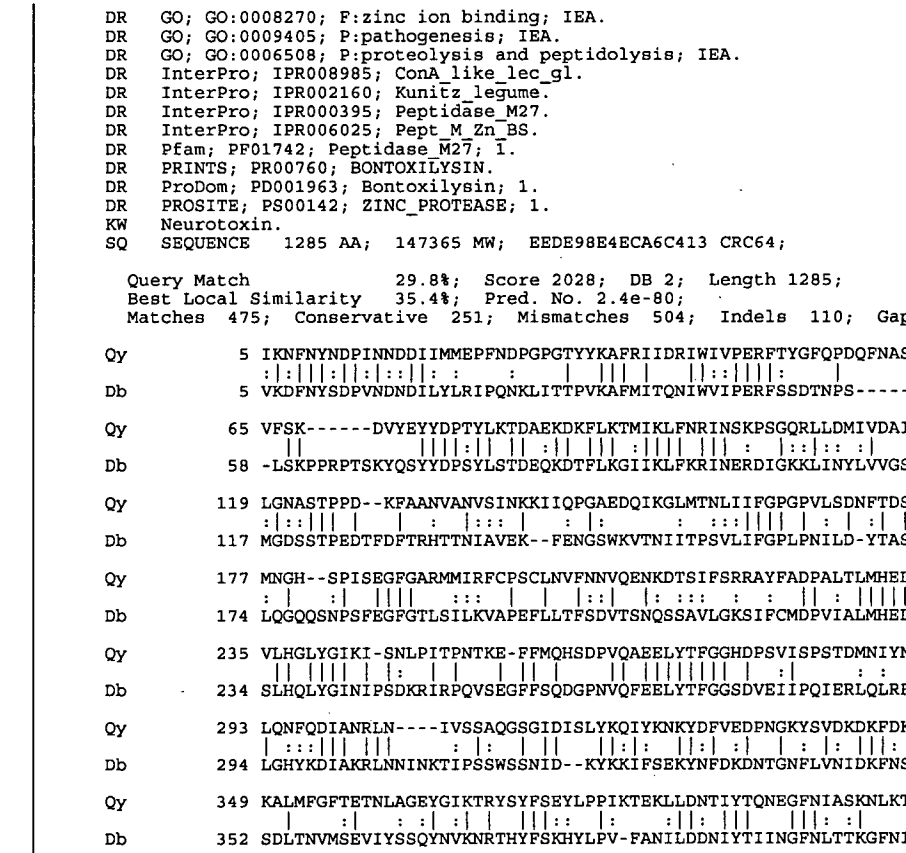
Qy        579 STNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNE 638

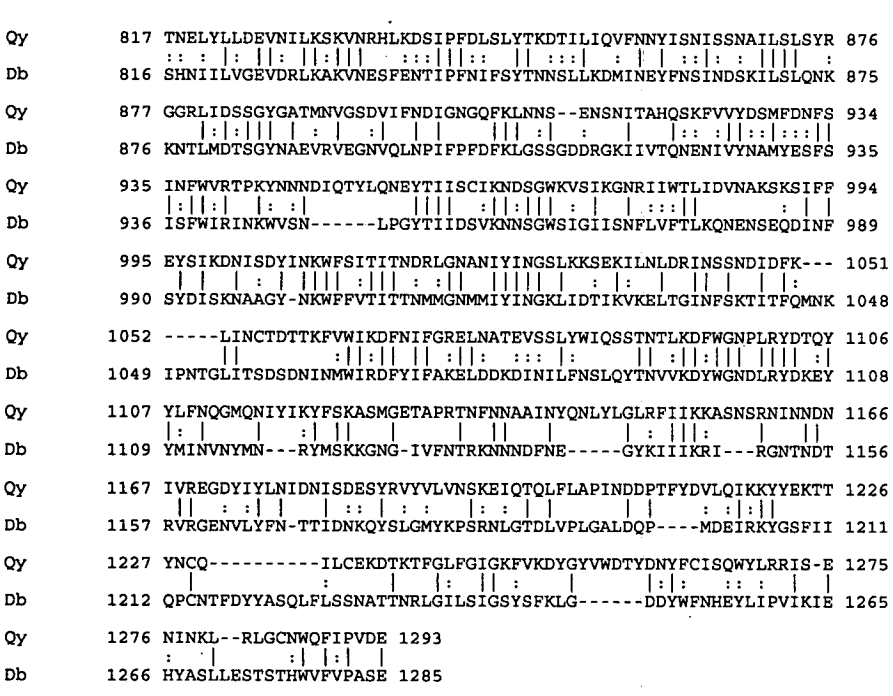
Db        579 -PSLAEKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNS 637
```



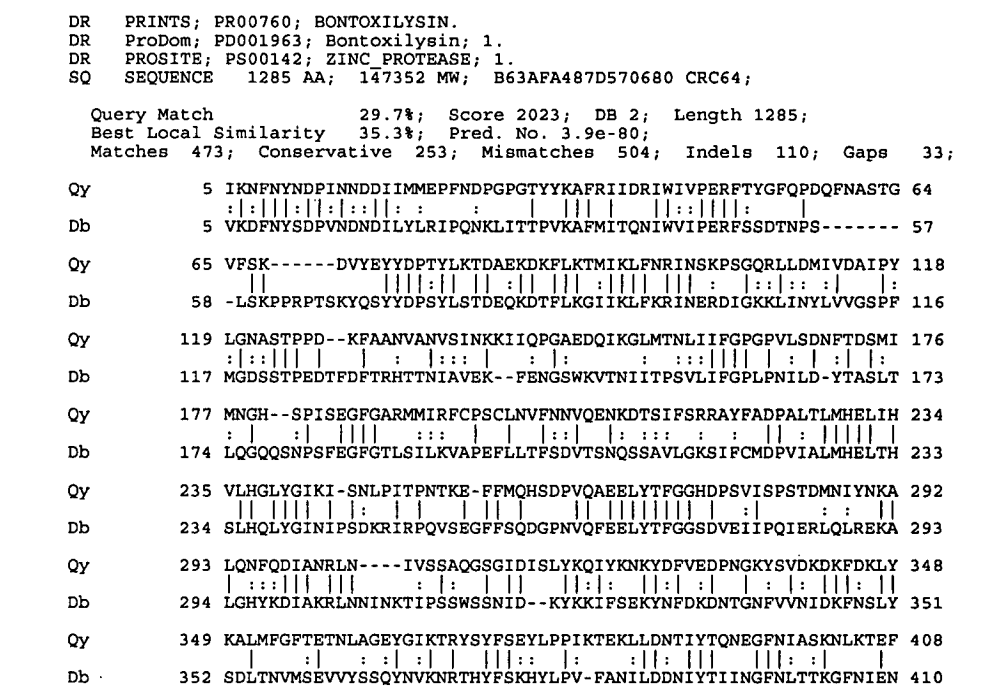
RESULT 14
Q45967
ID Q45967 PRELIMINARY; PRT; 1285 AA.
AC Q45967;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE D;
RX MEDLINE=96283801; PubMed=8679691;
RA Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA Ogumad K.;
RT "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT types C and D organisms.";
RL Biochim. Biophys. Acta 1307:123-126(1996).
DR EMBL; D38442; BAA07477.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1285 AA; 147365 MW; EEDE98E4ECA6C413 CRC64;

Query Match 29.8%; Score 2028; DB 2; Length 1285;
Best Local Similarity 35.4%; Pred. No. 2.4e-80;
Matches 475; Conservative 251; Mismatches 504; Indels 110; Gaps 33;





```
RESULT 15
Q9LBR1
ID   Q9LBR1      PRELIMINARY;      PRT;  1285 AA.
AC   Q9LBR1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   NT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-4947;
RA   Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA   Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT   "Characterization of the Progenitor Toxin Components Produced by
RT   Clostridium botulinum Type D Strain 4947.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037920; BAA90661.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
```



```
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1285 AA;  147352 MW;  B63AFA487D570680 CRC64;

  Query Match             29.7%;  Score 2023;  DB 2;  Length 1285;
  Best Local Similarity   35.3%;  Pred. No. 3.9e-80;
  Matches  473;  Conservative  253;  Mismatches  504;  Indels  110;  Gaps   33;
```

```
Qy       935 INFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFF 994
Db       936 ISFWIRINKWVSN------LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNENSEQDINF 989

Qy       995 EYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFK--- 1051
Db       990 SYDISKNAAGY-NKWFFVTITTNMMGNMMIYINGKLIDTIKVKELTGINFSKTITFQMNK 1048

Qy      1052 -----LINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQY 1106
Db      1049 IPNTGLITSDSDNINMWIRDFYIFAKELDDKDINILFNSLQYTNVVKDYWGNDLRYDKEY 1108

Qy      1107 YLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDN 1166
Db      1109 YMINVNYMN---RYMSKKGNG-IVFNTRKNNNDFNE-----GYKIIIKRI---RGNTNDT 1156

Qy      1167 IVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTT 1226
Db      1157 RVRGENVLYFN-TTIDNKQYSLGMYKPSRNLGTDLVPLGALDQP----MDEIRKYGSFII 1211

Qy      1227 YNCQ----------ILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRIS-E 1275
Db      1212 QPCNTFDYYASQLFLSSNATTNRLGILSIGSYSFKLG------DDYWFNHEYLIPVIKIE 1265

Qy      1276 NINKL--RLGCNWQFIPVDE 1293
Db      1266 HYASLLESTSTHWVFVPASE 1285
```

Search completed: September 10, 2004, 03:41:52
Job time : 64.3442 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:48:37 ; Search time 27.7139 Seconds
(without alignments)
4822.308 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 473
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	473	100.0	473	2	AAW68400	Aaw68400 Clostridi
2	447	94.5	449	3	AAY77139	Aay77139 Synthetic
3	447	94.5	449	4	AAB04167	Aab04167 Botulism
4	31	6.6	472	2	AAW68393	Aaw68393 Clostridi
5	31	6.6	472	2	AAW68394	Aaw68394 Clostridi
6	26	5.5	448	2	AAW68399	Aaw68399 Clostridi
7	26	5.5	451	2	AAW68395	Aaw68395 Clostridi
8	26	5.5	451	2	AAW68398	Aaw68398 Clostridi
9	26	5.5	452	2	AAW68396	Aaw68396 Clostridi
10	26	5.5	462	2	AAR95009	Aar95009 Type A ne
11	26	5.5	462	2	AAW68390	Aaw68390 Clostridi
12	26	5.5	462	2	AAW68397	Aaw68397 Clostridi
13	23	4.9	393	4	AAG68149	Aag68149 Streptomy
14	23	4.9	393	4	AAB74844	Aab74844 S. griseu
15	23	4.9	423	2	AAW68098	Aaw68098 Chlamydom
16	22	4.7	22	4	AAB67674	Aab67674 Amino aci
17	22	4.7	22	5	AAM49518	Aam49518 His-tag p
18	22	4.7	26	7	ADD31745	Add31745 Purificat
19	22	4.7	53	2	AAW36986	Aaw36986 HFKT pept
20	22	4.7	54	2	AAW36990	Aaw36990 HFKT pept
21	22	4.7	91	2	AAY09127	Aay09127 Recombina
22	22	4.7	93	2	AAY09128	Aay09128 Recombina
23	22	4.7	195	3	AAB11377	Aab11377 B. stearo
24	22	4.7	269	5	AAM49520	Aam49520 B. mori 1
25	22	4.7	269	5	AAM49541	Aam49541 B. mori T
26	22	4.7	323	2	AAW71185	Aaw71185 FP505 pro
27	22	4.7	332	4	AAE02191	Aae02191 Staphyloc
28	22	4.7	364	3	AAB13011	Aab13011 Q-tagged
29	22	4.7	567	4	AAU00220	Aau00220 Bad-DTTR
30	21	4.4	21	2	AAW23653	Aaw23653 Recombina
31	21	4.4	21	2	AAW23650	Aaw23650 Recombina
32	21	4.4	21	2	AAW23647	Aaw23647 Recombina
33	21	4.4	194	5	AAM51139	Aam51139 HIV TAT t
34	20	4.2	411	4	AAU00219	Aau00219 Bcl-Xl-DT
35	17	3.6	24	2	AAW19779	Aaw19779 Polyhisti
36	17	3.6	24	2	AAY06647	Aay06647 Polyhisti
37	17	3.6	24	2	AAY27015	Aay27015 Met Gly H
38	17	3.6	24	2	AAY09370	Aay09370 12 amino
39	17	3.6	24	3	AAY81959	Aay81959 Enterokin
40	17	3.6	24	4	AAB47057	Aab47057 Peptide f
41	17	3.6	24	6	ABG73858	Abg73858 Plasmid p
42	17	3.6	24	6	ABU10203	Abu10203 Peptide f
43	17	3.6	128	6	ABB99729	Abb99729 Amino aci
44	17	3.6	135	5	ABJ15133	Abj15133 Synthetic
45	17	3.6	135	5	ABJ15135	Abj15135 HisNeubla

ALIGNMENTS

RESULT 1
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30596.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC express BotG soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which

CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX
SQ Sequence 473 AA;

Query Match 100.0%; Score 473; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
```

RESULT 2
AAY77139
ID AAY77139 standard; protein; 449 AA.
XX
AC AAY77139;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype G (BoNTG) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype G; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87217.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 28; Page 47-48; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast
XX
SQ Sequence 449 AA;

Query Match 94.5%; Score 447; DB 3; Length 449;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 62

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 122

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 182

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 242

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 302

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 362

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 422

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||
Db        423 RRISENINKLRLGCNWQFIPVDEGWTE 449
```

```
RESULT 3
AAB04167
ID   AAB04167 standard; protein; 449 AA.
XX
AC   AAB04167;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype G).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54491.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 10b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 449 AA;

  Query Match             94.5%;  Score 447;  DB 4;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 62

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 122

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 182

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 242

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 302

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 362

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 422

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||
Db        423 RRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 4
AAW68393
ID   AAW68393 standard; protein; 472 AA.
XX
AC   AAW68393;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B strain Eklund 17B.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30580.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 300-302; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
```

CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 6.6%; Score 31; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 4.8e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31
||||||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31

RESULT 5
AAW68394
ID AAW68394 standard; protein; 472 AA.
XX
AC AAW68394;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B Danish strain.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30581.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 303-305; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 6.6%; Score 31; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 4.8e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31
||||||||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMASMADTILI 31

RESULT 6
AAW68399
ID AAW68399 standard; protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum; serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30593.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (202F strain) type F neurotoxin, encoded by a DNA
CC sequence (see AAV30593) in plasmid pETHisb. This vector can be used to
CC express BotC soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which
CC allow for the isolation of soluble recombinant proteins free of
CC significant endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of vaccines
CC and antitoxins that are useful in the treatment of humans and animals at
CC risk of intoxication with clostridial toxin
XX
SQ Sequence 448 AA;

Query Match 5.5%; Score 26; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
||||||||||||||||||||||||||

```
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 7
AAW68395
ID   AAW68395 standard; protein; 451 AA.
XX
AC   AAW68395;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
XX
OS   Clostridium botulinum; serotype E strain Belgua.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30584.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 324-325; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Belgua strain) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30584) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 8
AAW68398
ID   AAW68398 standard; protein; 451 AA.
XX
AC   AAW68398;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type D toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW   botulism; BotD.
XX
OS   Clostridium botulinum; serotype D.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30591.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 47; Page 352-353; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum seotype D neurotoxin, encoded by a DNA sequence
CC   (see AAV30591) in plasmid pETHisb. This vector can be used to express
CC   BotD soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on an affinity column. The
CC   invention relates to recombinant proteins derived from C. botulinum
CC   toxins, especially type B and type E toxins. Methods are provided which
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as immunogens for
CC   the production of vaccines and antitoxins useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin
XX
SQ   Sequence 451 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 9
AAW68396
ID   AAW68396 standard; protein; 452 AA.
XX
AC   AAW68396;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin E C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE.
```

```
XX
OS   Clostridium botulinum; serotype E strain NCTC 11219.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;    97WO-US015394.
XX
PR   28-AUG-1996;    96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 10
AAR95009
ID   AAR95009 standard; protein; 462 AA.
XX
AC   AAR95009;
XX
DT   07-JUL-1996  (first entry)
XX
DE   Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW   Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW   Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= Polyhistidine_tag
FT   Protein         22. .462
FT                   /label= C-fragment
XX
PN   WO9612802-A1.
XX
PD   02-MAY-1996.
XX
PF   23-OCT-1995;    95WO-US013737.
XX
PR   24-OCT-1994;    94US-00329154.
PR   16-MAR-1995;    95US-00405496.
PR   14-APR-1995;    95US-00422711.
PR   07-JUN-1995;    95US-00480604.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Padhye NV,  Kink JA,  Thalley BS,  Stafford DC;
PI   Firca JR;
XX
DR   WPI; 1996-230603/23.
DR   N-PSDB; AAT29246.
XX
PT   Fusion proteins comprising non-toxin protein and part of toxin - useful
PT   to form anti-toxins against Clostridium botulinum type A, and C.
PT   difficile type toxins, and to treat C. difficile intoxication, partic.
PT   diarrhoea.
XX
PS   Claim 7; Page 340-342; 434pp; English.
XX
CC   pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC   (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC   fragment C (see also AAR95008) of the Clostridium botulinum type A
CC   neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC   soluble protein and was purified by metal chelate affinity chromatography
CC   to obtain a product free of endotoxin contamination that may be useful as
CC   an immunogen in vaccine compsns
XX
SQ   Sequence 462 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 11
AAW68390
ID   AAW68390 standard; protein; 462 AA.
XX
AC   AAW68390;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW   Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW   botulism.
XX
OS   Clostridium botulinum; serotype A.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal histidine tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;    97WO-US015394.
XX
```

```
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30572, AAV30575.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 24; Page 265-267; 428pp; English.
XX
CC   This is the amino acid sequence of a histidine-tagged fragment C
CC   polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC   sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC   AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC   non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC   recombinant C fragment being purified on a poly-histidine binding
CC   affinity resin. The invention relates to recombinant proteins derived
CC   from C. botulinum toxins. Methods are provided which allow the isolation
CC   of soluble recombinant proteins that are free of significant endotoxin
CC   contamination. Preferred hosts for production of recombinant proteins are
CC   E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC   immunogens for the production of vaccines and antitoxins that are useful
CC   in the treatment of humans and animals at risk of intoxication with
CC   clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 12
AAW68397
ID   AAW68397 standard; protein; 462 AA.
XX
AC   AAW68397;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type C1 toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW   botulism; BotC.
XX
OS   Clostridium botulinum; serotype C1 strain Stockholm.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30588.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 45; Page 339-341; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Stockholm strain) type C1 neurotoxin, encoded by a
CC   DNA sequence (see AAV30588) in plasmid pETHisb. This vector is used to
CC   express BotC soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 462 AA;

  Query Match             5.5%;  Score 26;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-16;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 13
AAG68149
ID   AAG68149 standard; protein; 393 AA.
XX
AC   AAG68149;
XX
DT   24-JAN-2002  (first entry)
XX
DE   Streptomyces griseus RppA protein with a His tag SEQ ID NO:5.
XX
KW   Streptomyces griseus; RppA; pyran-2-one; drug; agricultural.
XX
OS   Streptomyces griseus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= His_tag
FT   Protein         22. .393
FT                   /label= RppA_protein
XX
PN   JP2001238677-A.
XX
PD   04-SEP-2001.
XX
PF   01-MAR-2000; 2000JP-00055782.
XX
PR   01-MAR-2000; 2000JP-00055782.
XX
PA   (MITR ) MITSUBISHI RAYON CO LTD.
XX
DR   WPI; 2001-613891/71.
XX
PT   Preparation of pyran-2-one derivative, used as intermediate and as
PT   additive for thermosensible transfer paper, comprises contacting carbonyl
PT   compound with microbe containing polyketide synthase gene.
XX
PS   Example 1; Page 10-11; 11pp; Japanese.
XX
CC   The present invention describes the preparation of pyran-2-one
```

CC derivatives. The preparation method comprises contacting carbonyl
CC compounds with microbes containing a polyketide synthase gene. Pyran-2-
CC one derivatives can be used as synthetic intermediates for drugs and
CC agricultural chemicals. They can also be used as additives for
CC thermosensible transfer paper. The present sequence represents the
CC Streptomyces griseus RppA protein with an N-terminal His tag, which is
CC used in an example from the present invention
XX
SQ Sequence 393 AA;

Query Match 4.9%; Score 23; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

RESULT 14
AAB74844
ID AAB74844 standard; protein; 393 AA.
XX
AC AAB74844;
XX
DT 18-JUN-2001 (first entry)
XX
DE S. griseus actinomycetes polyketide synthase with a histidine tag.
XX
KW Streptomyces griseus; actinomycete; polyketide synthase; rppA; drug;
KW 1,3,6,8-tetrahydroxynaphthalene; melanine; agricultural chemical;
KW malonyl CoA.
XX
OS Streptomyces griseus.
OS Synthetic.
XX
PN JP2000342269-A.
XX
PD 12-DEC-2000.
XX
PF 02-JUN-1999; 99JP-00155526.
XX
PR 02-JUN-1999; 99JP-00155526.
XX
PA (MITR) MITSUBISHI RAYON CO LTD.
XX
DR WPI; 2001-303158/32.
XX
PT Preparation of 1,3,6,8-tetrahydroxynaphthalene for use as a synthetic
PT material for melanine and as an intermediate for drugs and agricultural
PT chemicals, comprises using a microbe and an enzyme.
XX
PS Example 1; Page 9-10; 10pp; Japanese.
XX
CC The present invention describes a method for preparing 1,3,6,8-
CC tetrahydroxynaphthalene comprising using a microbe cell containing an
CC rppA gene encoding actinomycetes polyketide synthase, to convert malonyl
CC CoA as the substrate or a substance which can produce malonyl CoA in the
CC microbe cell to 1,3,6,8-tetrahydroxynaphthalene and recovering it.
CC 1,3,6,8-tetrahydroxynaphthalene is used as a synthetic material for
CC melanine and as an intermediate for drugs and agricultural chemicals. The
CC present sequence represents the actinomycetes polyketide synthase protein
CC isolated from Streptomyces griseus comprising an histidine tag, which is
CC used in an example from the present invention
XX
SQ Sequence 393 AA;

Query Match 4.9%; Score 23; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

RESULT 15
AAW68098
ID AAW68098 standard; protein; 423 AA.
XX
AC AAW68098;
XX
DT 27-OCT-1998 (first entry)
XX
DE Chlamydomonas reinhardtii histidine tagged RB47 binding protein.
XX
KW RB47; activator; mRNA binding protein; psbA; chloroplast; recombinant;
KW regulated production; protein synthesis; plant propagation.
XX
OS Chlamydomonas reinhardtii.
OS Synthetic.
XX
PN WO9831823-A1.
XX
PD 23-JUL-1998.
XX
PF 16-JAN-1998; 98WO-US000840.
XX
PR 17-JAN-1997; 97US-0035955P.
PR 12-DEC-1997; 97US-0069400P.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Mayfield S;
XX
DR WPI; 1998-427562/36.
DR N-PSDB; AAV41263.
XX
PT New isolated RB47 and RB60 genes - used for, e.g. regulated expression of
PT desired recombinant molecule in prokaryote and eukaryote, particularly
PT plant cells.
XX
PS Claim 10; Fig 5; 109pp; English.
XX
CC The sequence is that of a histidine tagged RB47 mRNA binding protein. It
CC can be used in the regulated production of a desired recombinant
CC molecule. It can be used as a cell culture system, an in vitro expression
CC system or a whole tissue, preferably a plant and allows for a high level
CC of protein synthesis in plant chloroplasts and in the cytoplasm of both
CC prokaryotic and eukaryotic cells. It can also be used to screen for
CC agonists or antagonists of the binding of RB47 to the RB47 binding site.
CC Antagonists of the binding are useful in the prevention of plant
CC propagation. They can also be used for screening for agonists or
CC antagonists of RB60. Such agonists or antagonists would be useful in
CC general to modify expression of RB60 as a way to regulate cellular
CC processes in a redox manner
XX
SQ Sequence 423 AA;

Query Match 4.9%; Score 23; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.4e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGHHHHHHHHHHSSGHIEGRHMA 23
||||||||||||||||||||||||
Db 1 MGHHHHHHHHHHSSGHIEGRHMA 23

Search completed: September 10, 2004, 05:06:05
Job time : 28.7139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:58:34 ; Search time 7.75508 Seconds
(without alignments)
3148.784 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 473
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	5.5	462	1	US-08-480-604A-26	Sequence 26, Appl
2	26	5.5	462	2	US-08-405-496A-26	Sequence 26, Appl
3	26	5.5	462	3	US-08-915-136-26	Sequence 26, Appl
4	26	5.5	462	4	US-09-084-517-26	Sequence 26, Appl
5	23	4.9	23	1	US-08-480-604A-24	Sequence 24, Appl
6	23	4.9	23	2	US-08-405-496A-24	Sequence 24, Appl
7	23	4.9	23	3	US-08-915-136-24	Sequence 24, Appl
8	23	4.9	23	4	US-09-084-517-24	Sequence 24, Appl
9	22	4.7	22	4	US-09-660-742-3	Sequence 3, Appli
10	22	4.7	53	2	US-08-651-818A-19	Sequence 19, Appl
11	22	4.7	53	3	US-09-184-826-19	Sequence 19, Appl
12	22	4.7	54	2	US-08-651-818A-23	Sequence 23, Appl
13	22	4.7	54	3	US-09-184-826-23	Sequence 23, Appl
14	22	4.7	323	1	US-08-591-989-5	Sequence 5, Appli
15	21	4.4	21	2	US-08-651-818A-21	Sequence 21, Appl
16	21	4.4	21	3	US-09-184-826-21	Sequence 21, Appl
17	17	3.6	24	3	US-08-584-031-8	Sequence 8, Appli
18	17	3.6	24	3	US-08-780-496-8	Sequence 8, Appli
19	17	3.6	24	4	US-08-928-069-9	Sequence 9, Appli
20	15	3.2	20	4	US-09-186-489-6	Sequence 6, Appli
21	15	3.2	20	4	US-10-043-665B-6	Sequence 6, Appli
22	12	2.5	16	5	PCT-US94-04361-56	Sequence 56, Appl
23	12	2.5	17	2	US-08-600-783-13	Sequence 13, Appl
24	11	2.3	20	4	US-09-674-677-34	Sequence 34, Appl
25	11	2.3	24	4	US-09-292-858B-27	Sequence 27, Appl
26	11	2.3	254	3	US-09-167-434-7	Sequence 7, Appli
27	11	2.3	254	3	US-08-853-755-7	Sequence 7, Appli
28	11	2.3	302	2	US-08-203-532F-4	Sequence 4, Appli
29	11	2.3	302	3	US-08-950-860-16	Sequence 16, Appl
30	11	2.3	302	3	US-09-078-465-4	Sequence 4, Appli
31	11	2.3	302	5	PCT-US95-01882A-4	Sequence 4, Appli
32	11	2.3	303	2	US-08-203-532F-2	Sequence 2, Appli
33	11	2.3	303	3	US-09-078-465-2	Sequence 2, Appli
34	11	2.3	303	5	PCT-US95-01882A-2	Sequence 2, Appli
35	11	2.3	363	4	US-09-328-352-4930	Sequence 4930, Ap
36	11	2.3	414	5	PCT-US92-06840-2	Sequence 2, Appli
37	11	2.3	452	1	US-07-618-312A-2	Sequence 2, Appli
38	11	2.3	452	1	US-07-618-312A-4	Sequence 4, Appli
39	11	2.3	452	1	US-08-110-786A-8	Sequence 8, Appli
40	11	2.3	452	1	US-08-280-228-2	Sequence 2, Appli
41	11	2.3	452	1	US-08-280-228-4	Sequence 4, Appli
42	11	2.3	618	1	US-08-668-381A-5	Sequence 5, Appli
43	11	2.3	853	4	US-08-913-880C-17	Sequence 17, Appl
44	11	2.3	858	4	US-08-913-880C-16	Sequence 16, Appl
45	11	2.3	860	4	US-08-913-880C-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             5.5%;  Score 26;  DB 1;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match             5.5%;  Score 26;  DB 2;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             5.5%;  Score 26;  DB 3;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-17;
```

```
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-26

  Query Match             5.5%;  Score 26;  DB 4;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26
              ||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMA 26


RESULT 5
US-08-480-604A-24
; Sequence 24, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-24

  Query Match             4.9%;  Score 23;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 6
US-08-405-496A-24
```

```
; Sequence 24, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
;   APPLICANT: WILLIAMS, JAMES A.
;   TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
;   TITLE OF INVENTION: NEUROTOXIN
;   NUMBER OF SEQUENCES: 30
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: MEDLEN & CARROLL, LLP
;     STREET: 220 MONTGOMERY STREET, SUITE 2200
;     CITY: SAN FRANCISCO
;     STATE: CALIFORNIA
;     COUNTRY: USA
;     ZIP: 94104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/405,496A
;     FILING DATE: 16-MAR-1995
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/329,154
;     FILING DATE: 25-OCT-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/161,907
;     FILING DATE: 02-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/985,321
;     FILING DATE: 04-DEC-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/429,791
;     FILING DATE: 31-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: INGOLIA, DIANE E.
;     REGISTRATION NUMBER: 40,027
;     REFERENCE/DOCKET NUMBER: OPHD-01308
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 705-8410
;     TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 24:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 23 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: unknown
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-405-496A-24

  Query Match             4.9%;  Score 23;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 7
US-08-915-136-24
; Sequence 24, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
;   APPLICANT: KINK, JOHN A.
;   APPLICANT: THALLEY, BRUCE S.
;   APPLICANT: PADHYE, NISHA V.
;   APPLICANT: FIRCA, JOSEPH R.
;   APPLICANT: STAFFORD, DOUGLAS C.
;   TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
;   TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
;   NUMBER OF SEQUENCES: 32
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: MEDLEN & CARROLL, LLP
;     STREET: 220 MONTGOMERY STREET, SUITE 2200
;     CITY: SAN FRANCISCO
;     STATE: CALIFORNIA
;     COUNTRY: UNITED STATES OF AMERICA
;     ZIP: 94104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/915,136
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/480,604
;     FILING DATE:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/405,496
;     FILING DATE: 16-MAR-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/329,154
;     FILING DATE: 25-OCT-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/161,907
;     FILING DATE: 02-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/985,321
;     FILING DATE: 04-DEC-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/429,791
;     FILING DATE: 31-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: INGOLIA, DIANE E.
;     REGISTRATION NUMBER: 40,027
;     REFERENCE/DOCKET NUMBER: OPHD-01763
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 705-8410
;     TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 24:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 23 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: unknown
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-915-136-24

  Query Match             4.9%;  Score 23;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 8
US-09-084-517-24
; Sequence 24, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
;   APPLICANT: KINK, JOHN A.
;   APPLICANT: WILLIAMS, JAMES A.
;   TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
;   TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
;   NUMBER OF SEQUENCES: 30
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
;     STREET: 220 MONTGOMERY STREET, SUITE 2200
;     CITY: SAN FRANCISCO
```

```
;       STATE:  CALIFORNIA
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/084,517
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/
;       FILING DATE:  16-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/329,154
;       FILING DATE:  25-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/161,907
;       FILING DATE:  02-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/985,321
;       FILING DATE:  04-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/429,791
;       FILING DATE:  31-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CARROLL, PETER G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  OPHD-01610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  23 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-084-517-24

  Query Match             4.9%;  Score 23;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMA 23
              |||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMA 23


RESULT 9
US-09-660-742-3
; Sequence 3, Application US/09660742
; Patent No. 6485977
; GENERAL INFORMATION:
;  APPLICANT:  Collmer, Alan
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  RECOMBINANT CONSTRUCTS AND TECHNIQUES FOR DELIVERING TO
;  TITLE OF INVENTION:  EUCARYOTIC CELLS BACTERIAL PROTEINS THAT ARE SECRETED
;  TITLE OF INVENTION:  VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE:  19603/4021
;  CURRENT APPLICATION NUMBER: US/09/660,742
;  CURRENT FILING DATE:  2000-09-13
;  PRIOR APPLICATION NUMBER:  60/153,507
;  PRIOR FILING DATE:  1999-09-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 22
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Polypeptide
US-09-660-742-3

  Query Match             4.7%;  Score 22;  DB 4;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 10
US-08-651-818A-19
; Sequence 19, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-19

  Query Match             4.7%;  Score 22;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 11
US-09-184-826-19
; Sequence 19, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
```

```
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-19

  Query Match             4.7%;  Score 22;  DB 3;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 12
US-08-651-818A-23
; Sequence 23, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-23

  Query Match             4.7%;  Score 22;  DB 2;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 13
US-09-184-826-23
; Sequence 23, Application US/09184826
; Patent No. 6248543
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/184,826
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-09-184-826-23

  Query Match             4.7%;  Score 22;  DB 3;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHM 22


RESULT 14
US-08-591-989-5
```

```
; Sequence 5, Application US/08591989
; Patent No. 5795721
;  GENERAL INFORMATION:
;    APPLICANT:  Ross S. Rabin, Sumedha Jayasena
;    APPLICANT:  and Larry Gold
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC
;    TITLE OF INVENTION:  ACID LIGANDS OF ICP4
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 East Prentice Avenue, Suite #200
;      CITY:  Englewood
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.40 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX 49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-591-989-5

  Query Match             4.7%;  Score 22;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 4e-13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHM 22
              ||||||||||||||||||||||
Db         17 MGHHHHHHHHHHSSGHIEGRHM 38


RESULT 15
US-08-651-818A-21
; Sequence 21, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-651-818A-21

  Query Match             4.4%;  Score 21;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-13;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRH 21
              |||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRH 21


Search completed: September 10, 2004, 05:17:11
Job time : 8.75508 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 05:07:28 ; Search time 23.9265 Seconds
(without alignments)
6339.663 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 473
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1335176 seqs, 320689617 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	473	100.0	473	12	US-10-272-898-79	Sequence 79, Appl
2	473	100.0	473	15	US-10-354-774-79	Sequence 79, Appl
3	473	100.0	473	15	US-10-271-012-79	Sequence 79, Appl
4	473	100.0	473	16	US-10-729-122-79	Sequence 79, Appl
5	473	100.0	473	16	US-10-729-039-79	Sequence 79, Appl
6	447	94.5	449	10	US-09-910-186A-18	Sequence 18, Appl
7	447	94.5	1297	12	US-10-272-898-77	Sequence 77, Appl
8	447	94.5	1297	15	US-10-354-774-77	Sequence 77, Appl
9	447	94.5	1297	15	US-10-271-012-77	Sequence 77, Appl
10	447	94.5	1297	15	US-10-452-024-7	Sequence 7, Appli
11	447	94.5	1297	15	US-10-452-024-150	Sequence 150, App
12	447	94.5	1297	15	US-10-205-516-14	Sequence 14, Appl
13	447	94.5	1297	16	US-10-729-122-77	Sequence 77, Appl
14	447	94.5	1297	16	US-10-729-039-77	Sequence 77, Appl
15	447	94.5	1307	15	US-10-205-516-28	Sequence 28, Appl
16	447	94.5	2017	15	US-10-452-024-146	Sequence 146, App
17	31	6.6	472	12	US-10-272-898-44	Sequence 44, Appl
18	31	6.6	472	12	US-10-272-898-46	Sequence 46, Appl
19	31	6.6	472	15	US-10-354-774-44	Sequence 44, Appl
20	31	6.6	472	15	US-10-354-774-46	Sequence 46, Appl
21	31	6.6	472	15	US-10-271-012-44	Sequence 44, Appl
22	31	6.6	472	15	US-10-271-012-46	Sequence 46, Appl
23	31	6.6	472	16	US-10-729-122-44	Sequence 44, Appl
24	31	6.6	472	16	US-10-729-122-46	Sequence 46, Appl
25	31	6.6	472	16	US-10-729-039-44	Sequence 44, Appl
26	31	6.6	472	16	US-10-729-039-46	Sequence 46, Appl
27	26	5.5	448	12	US-10-272-898-73	Sequence 73, Appl
28	26	5.5	448	15	US-10-354-774-73	Sequence 73, Appl
29	26	5.5	448	15	US-10-271-012-73	Sequence 73, Appl
30	26	5.5	448	16	US-10-729-122-73	Sequence 73, Appl
31	26	5.5	448	16	US-10-729-039-73	Sequence 73, Appl
32	26	5.5	451	12	US-10-272-898-54	Sequence 54, Appl
33	26	5.5	451	12	US-10-272-898-68	Sequence 68, Appl
34	26	5.5	451	15	US-10-354-774-54	Sequence 54, Appl
35	26	5.5	451	15	US-10-354-774-68	Sequence 68, Appl
36	26	5.5	451	15	US-10-271-012-54	Sequence 54, Appl
37	26	5.5	451	15	US-10-271-012-68	Sequence 68, Appl
38	26	5.5	451	16	US-10-729-122-54	Sequence 54, Appl
39	26	5.5	451	16	US-10-729-122-68	Sequence 68, Appl
40	26	5.5	451	16	US-10-729-039-54	Sequence 54, Appl
41	26	5.5	451	16	US-10-729-039-68	Sequence 68, Appl
42	26	5.5	452	12	US-10-272-898-56	Sequence 56, Appl
43	26	5.5	452	15	US-10-354-774-56	Sequence 56, Appl
44	26	5.5	452	15	US-10-271-012-56	Sequence 56, Appl
45	26	5.5	452	16	US-10-729-122-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-10-272-898-79
; Sequence 79, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:

```
;            LENGTH: 473 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-272-898-79

  Query Match            100.0%;  Score 473;  DB 12;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 2
US-10-354-774-79
; Sequence 79, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-354-774-79

  Query Match            100.0%;  Score 473;  DB 15;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 3
US-10-271-012-79
; Sequence 79, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-271-012-79

  Query Match           100.0%;  Score 473;  DB 15;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 4
US-10-729-122-79
; Sequence 79, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-729-122-79

  Query Match           100.0%;  Score 473;  DB 16;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 5
US-10-729-039-79
```

```
; Sequence 79, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-729-039-79

  Query Match           100.0%;  Score 473;  DB 16;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 6
US-09-910-186A-18
; Sequence 18, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 449
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-09-910-186A-18

  Query Match           94.5%;  Score 447;  DB 10;  Length 449;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 62

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 122

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 182

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 242

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 302

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 362
```

```
Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 422

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db        423 RRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 7
US-10-272-898-77
; Sequence 77, Application US/10272898
; Publication No. US20030162745A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/272,898
;             FILING DATE: 04-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-272-898-77

  Query Match             94.5%;  Score 447;  DB 12;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 8
US-10-354-774-77
; Sequence 77, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-354-774-77

  Query Match             94.5%;  Score 447;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
                 ||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 9
US-10-271-012-77
; Sequence 77, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-271-012-77

  Query Match           94.5%;  Score 447;  DB 15;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
                 ||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 10
US-10-452-024-7
; Sequence 7, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-7

  Query Match           94.5%;  Score 447;  DB 15;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 11
US-10-452-024-150
; Sequence 150, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
;  APPLICANT: Simpson, Lance
;  APPLICANT:  Park, Jung-Beak
;  APPLICANT:  Maksymowych, Andrew
;  TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
;  FILE REFERENCE: 9855-96U1
;  CURRENT APPLICATION NUMBER: US/10/452,024
;  CURRENT FILING DATE:  2003-06-02
;  PRIOR APPLICATION NUMBER: 60/384,949
;  PRIOR FILING DATE: 2002-05-31
;  NUMBER OF SEQ ID NOS: 188
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 150
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-150

  Query Match             94.5%;  Score 447;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
               ||||||||||||||||||||||||||||
Db        1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 12
US-10-205-516-14
; Sequence 14, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 1297
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7)
;   OTHER INFORMATION: mixed sequence
US-10-205-516-14

  Query Match             94.5%;  Score 447;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
```

```
                    ||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 13
US-10-729-122-77
; Sequence 77, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-122-77

  Query Match             94.5%;  Score 447;  DB 16;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYYVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYYVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
               ||||||||||||||||||||||||||
Db        1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 14
US-10-729-039-77
; Sequence 77, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-039-77

  Query Match             94.5%;  Score 447;  DB 16;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
```

```
Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

```
RESULT 15
US-10-205-516-28
; Sequence 28, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
;  APPLICANT: Zhong, Jun
;  TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
;  TITLE OF INVENTION:  Neurotoxins Through Recombinant DNA Technique
;  FILE REFERENCE: JZbtx1
;  CURRENT APPLICATION NUMBER: US/10/205,516
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
;   LENGTH: 1307
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7)
;   OTHER INFORMATION: mixed sequence
US-10-205-516-28
```

```
  Query Match            94.5%;  Score 447;  DB 15;  Length 1307;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 914

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 974

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1034

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1094

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1154

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1214

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1274

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1275 RRISENINKLRLGCNWQFIPVDEGWTE 1301
```

```
Search completed: September 10, 2004, 05:53:05
Job time : 25.9265 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:33:02 ; Search time 6.55275 Seconds
(without alignments)
6943.440 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 473
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	447	94.5	1297	2	S39791	neurotoxin - Clost
2	11	2.3	302	2	A55641	homeotic protein G
3	11	2.3	303	2	A56837	homeotic protein M
4	11	2.3	303	2	B49122	homeobox protein M
5	11	2.3	303	2	A48130	growth arrest-spec
6	11	2.3	354	2	S39406	homeotic protein o
7	11	2.3	355	2	S35345	otx1 protein - mou
8	11	2.3	355	2	I56547	homeodomain protei
9	11	2.3	414	2	A40350	transcription repr
10	11	2.3	414	2	A48273	delta/YY1/NF-E1/UC
11	11	2.3	420	2	T39712	hypothetical prote
12	11	2.3	427	2	T42516	hypothetical prote
13	11	2.3	1180	2	S69205	stripe a/b protein
14	11	2.3	1315	1	BTCLTN	tentoxilysin (EC 3
15	10	2.1	132	2	I39193	gene HOXA1 protein
16	10	2.1	133	2	B30242	stem cell protein
17	10	2.1	222	2	I39192	gene HOXA1 protein
18	10	2.1	236	2	S41512	Brn-3b protein - m
19	10	2.1	298	2	S41469	homeotic protein M
20	10	2.1	309	2	T29293	hypothetical prote
21	10	2.1	321	2	T02987	myb-related protei
22	10	2.1	326	2	D83483	probable metal tra
23	10	2.1	331	2	A30242	homeotic protein E
24	10	2.1	335	2	G01448	homeobox protein H
25	10	2.1	351	1	KGZQHL	histidine-rich gly
26	10	2.1	375	2	A46390	cAMP receptor subt
27	10	2.1	387	2	A47446	HNF-3/fork head fa
28	10	2.1	390	1	A38565	polycomb (Pc) prot
29	10	2.1	391	2	H86187	hypothetical prote
30	10	2.1	407	2	G84783	probable pectinest
31	10	2.1	410	2	I38502	gene Brn-3b protei
32	10	2.1	411	2	I58156	Brn-3.2 - mouse
33	10	2.1	430	2	S66671	neuron-derived rec
34	10	2.1	451	2	A55909	transforming prote
35	10	2.1	469	2	I37451	HBF-G2 (HFK-2) pro
36	10	2.1	476	2	A54743	transcription fact
37	10	2.1	480	2	JH0672	brain factor 1 pro
38	10	2.1	580	2	T46024	hypothetical prote
39	10	2.1	623	1	S33167	gene pointed prote
40	10	2.1	625	2	S71930	neuron-derived rec
41	10	2.1	628	2	JC2493	neuron derived orp
42	10	2.1	633	1	A26030	serine/threonine-s
43	10	2.1	657	2	A29454	knob-associated hi
44	10	2.1	665	2	S02358	lamin A - African
45	10	2.1	717	2	S38177	SSV7 protein homol

ALIGNMENTS

RESULT 1
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 94.5%; Score 447; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
               |||||||||||||||||||||||||||
```

```
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 2
A55641
homeotic protein GAX - human
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 24-Sep-1999
C;Accession: A55641
R;LePage, D.F.; Altomare, D.A.; Testa, J.R.; Walsh, K.
Genomics 24, 535-540, 1994
A;Title: Molecular cloning and localization of the human GAX gene to 7p21.
A;Reference number: A55641; MUID:95229154; PMID:7713505
A;Accession: A55641
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <LEP>
A;Cross-references: GB:L36328; NID:g557549; PIDN:AAA58497.1; PID:g557550
C;Genetics:
A;Gene: GAX
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;186-242/Domain: homeobox homology <HOX>

```
  Query Match             2.3%;  Score 11;  DB 2;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 0.0031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77
```

RESULT 3
A56837
homeotic protein MOX2 - human
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 24-Sep-1999
C;Accession: A56837; S52168
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.S.; Theodorakis, K.; Mankoo, B.; Pachnis, V.;
Genomics 26, 550-555, 1995
A;Title: Isolation of the human MOX2 homeobox gene and localization to chromosome 7p22.1
A;Reference number: A56837; MUID:95331791; PMID:7607679
A;Accession: A56837
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-303 <GRI>
A;Cross-references: EMBL:X82629; NID:g732790; PIDN:CAA57949.1; PID:g732791
R;Grigoriou, M.; Kastrinaki, M.C.; Modi, W.; Theodorakis, K.; Maukoo, B.; Pachnis, V.; K
submitted to the EMBL Data Library, November 1994
A;Description: Isolation of the human MOX2 homeobox gene and localization to chromosome
A;Reference number: S52168
A;Accession: S52168
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-7,'R',9-110,'R',112-122,'Q',124-157,'V',159-303 <GR2>
A;Cross-references: EMBL:X82629
C;Genetics:
A;Gene: GDB:MOX2
A;Cross-references: GDB:120703; OMIM:155970
A;Map position: 3q13-3q13
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

```
  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77
```

RESULT 4
B49122
homeobox protein Mox-2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Dec-1993 #sequence_revision 17-Nov-1995 #text_change 17-Nov-2000
C;Accession: B49122; S41779; S29902
R;Candia, A.F.; Hu, J.; Crosby, J.; Lalley, P.A.; Noden, D.; Nadeau, J.H.; Wright, C.V.
Development 116, 1123-1136, 1992
A;Title: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially ex
A;Reference number: A49122; MUID:93201999; PMID:1363541
A;Accession: B49122
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN1>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
A;Experimental source: C57BL
A;Note: the complete translation is not shown
R;Candia, A.F.; Kovalik, J.P.; Wright, C.V.E.
Nucleic Acids Res. 21, 4982, 1993
A;Title: Amino acid sequence of Mox-2 and comparison to its Xenopus and rat homologs.
A;Reference number: S41469; MUID:94232829; PMID:7909944
A;Accession: S41779
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-303 <CAN2>
A;Cross-references: EMBL:Z16406; NID:g57949; PIDN:CAA78899.1; PID:g57950
R;Candia, A.
submitted to the EMBL Data Library, October 1992
A;Reference number: S29902
A;Accession: S29902
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 186-246 <CAN>
A;Cross-references: EMBL:Z16406
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

```
  Query Match             2.3%;  Score 11;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.0031;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77
```

RESULT 5
A48130
growth arrest-specific homeobox protein Cvx - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A48130; S31976
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
Mol. Cell. Biol. 13, 3722-3733, 1993
A;Title: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated dur
A;Reference number: A48130; MUID:93268321; PMID:8098844
A;Accession: A48130
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-303 <GOR>
A;Cross-references: GB:Z17223; NID:g57951; PIDN:CAA78931.1; PID:g57952
A;Experimental source: aorta
A;Note: sequence extracted from NCBI backbone (NCBIN:132842, NCBIP:132843)
R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
submitted to the EMBL Data Library, October 1992
A;Description: Cvx: A diverged homeodomain gene which is rapidly down-regulated following
A;Reference number: S31976
A;Accession: S31976
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-182,'S',184-303 <GO2>

A;Cross-references: EMBL:Z17223
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;187-243/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 67 GHHHHHHHHHH 77

RESULT 6
S39406
homeotic protein otx1 - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 15-Oct-1999
C;Accession: S39406
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; E
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S39406
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-354 <SIM>
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 7
S35345
otx1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000
C;Accession: S35345; S29087
R;Simeone, A.; Acampora, D.; Mallamaci, A.; Stornaiuolo, A.; d'Apice, M.R.; Nigro, V.; E
EMBO J. 12, 2735-2747, 1993
A;Title: A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid
A;Reference number: S35345; MUID:93327763; PMID:8101484
A;Accession: S35345
A;Molecule type: DNA
A;Residues: 1-355 <SIM>
R;Simeone, A.; Acampora, D.; Gulisano, M.; Stornaiuolo, A.; Boncinelli, E.
Nature 358, 687-690, 1992
A;Title: Nested expression domains of four homeobox genes in developing rostral brain.
A;Reference number: S29084; MUID:92357147; PMID:1353865
A;Accession: S29087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 31-98 <SI2>
A;Cross-references: EMBL:X68883; NID:g53540; PIDN:CAA48754.1; PID:g53541
C;Genetics:
A;Gene: otx1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 8
I56547
homeodomain protein otx1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C;Accession: I56547
R;Frantz, G.D.; Weimann, J.M.; Levin, M.E.; McConnell, S.K.
J. Neurosci. 14, 5725-5740, 1994
A;Title: Otx1 and Otx2 define layers and regions in developing cerebral cortex and cereb
A;Reference number: I56547; MUID:95016961; PMID:7931541
A;Accession: I56547
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: GB:L32602; NID:g535739; PIDN:AAA53557.1; PID:g535740
C;Genetics:
A;Gene: OTX1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;39-95/Domain: homeobox homology <HOX>

Query Match 2.3%; Score 11; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GHHHHHHHHHH 12
||||||||||
Db 291 GHHHHHHHHHH 301

RESULT 9
A40350
transcription repressor protein YY1 - human
N;Alternate names: transcription repressor/activator protein NF-E1
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Oct-1999
C;Accession: A40350; S78494; S33712; A56419
R;Shi, Y.; Seto, E.; Chang, L.S.; Shenk, T.
Cell 67, 377-388, 1991
A;Title: Transcriptional repression by YY1, a human GLI-Krueppel-related protein, and re
A;Reference number: A40350; MUID:92005716; PMID:1655281
A;Accession: A40350
A;Molecule type: mRNA
A;Residues: 1-414 <SHI>
A;Cross-references: GB:M77698; NID:g186767; PIDN:AAA59467.1; PID:g186768
A;Experimental source: HeLa cells
A;Note: the authors translated the codon CGC for residue 371 as Lys, CGA for residue 375
R;Whitson, R.H.; Huang, T.; Dang, J.; Itakura, K.
submitted to the EMBL Data Library, July 1992
A;Description: Observed and predicted DNA binding of a zinc finger protein which recogni
A;Reference number: S78494
A;Accession: S78494
A;Molecule type: mRNA
A;Residues: 1-195,'G',197-414 <WHI>
A;Cross-references: EMBL:Z14077; NID:g38010; PIDN:CAA78455.1; PID:g38011
R;Park, K.; Atchison, M.L.
Proc. Natl. Acad. Sci. U.S.A. 88, 9804-9808, 1991
A;Title: Isolation of a candidate repressor/activator, NF-E1 (YY-1), delta, that binds t
A;Reference number: A56419; MUID:92052179; PMID:1946405
A;Accession: S33712
A;Molecule type: mRNA
A;Residues: 1-64,'R',66-195,'G',197-414 <PAR>
A;Cross-references: GB:M76541; NID:g189173; PIDN:AAA59926.1; PID:g189174
C;Keywords: DNA binding; transcription regulation; zinc finger
F;298-320/Region: zinc finger CCHH motif
F;327-347/Region: zinc finger CCHH motif

```
F;355-377/Region: zinc finger CCHH motif
F;385-407/Region: zinc finger CCHH motif

  Query Match             2.3%;  Score 11;  DB 2;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0041;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79


RESULT 10
A48273
delta/YY1/NF-E1/UCRBP transcription factor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 01-Dec-2000
C;Accession: A48273; A42055; A56418
R;Safrany, G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 5559-5563, 1993
A;Title: Characterization of the mouse gene that encodes the delta/YY1/NF-E1/UCRBP trans
A;Reference number: A48273; MUID:93296177; PMID:8516301
A;Accession: A48273
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-414 <RES>
A;Cross-references: GB:L13968; NID:g293847; PIDN:AAA40477.1; PID:g293849
R;Flanagan, J.R.; Becker, K.G.; Ennist, D.L.; Gleason, S.L.; Driggers, P.H.; Levi, B.Z.;
Mol. Cell. Biol. 12, 38-44, 1992
A;Title: Cloning of a negative transcription factor that binds to the upstream conserved
A;Reference number: A42055; MUID:92107191; PMID:1309593
A;Accession: A42055
A;Molecule type: mRNA
A;Residues: 1-414 <FLA>
A;Cross-references: GB:M73963; NID:g202270; PIDN:AAA40522.1; PID:g202271
A;Note: sequence extracted from NCBI backbone (NCBIN:74641, NCBIP:74642)
R;Hariharan, N.; Kelley, D.E.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 88, 9799-9803, 1991
A;Title: delta, a transcription factor that binds to downstream elements in several poly
A;Reference number: A56418; MUID:92052178; PMID:1946404
A;Accession: A56418
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-218,'S',220-374,'G',376-414 <HAR>
A;Cross-references: GB:M74590; NID:g192940; PIDN:AAA37521.1; PID:g192941
C;Genetics:
A;Introns: 227/1; 281/2; 301/3; 354/3
C;Keywords: transcription factor; zinc finger

  Query Match             2.3%;  Score 11;  DB 2;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0041;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 11
T39712
hypothetical protein SPBC17D11.01 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: hypothetical protein SPBC20F10.11
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Jan-2000
C;Accession: T39712; T39882
R;Beck, A.; Reinhardt, R.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21872
A;Accession: T39712
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-420 <BEC>
A;Cross-references: EMBL:AL031739; PIDN:CAA21072.1; GSPDB:GN00067; SPDB:SPBC17D11.01
A;Experimental source: strain 972h-; cosmid c17D11
R;Lyne, M.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21888
A;Accession: T39882
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-281 <LYN>
A;Cross-references: EMBL:AL021747; PIDN:CAA16851.1; GSPDB:GN00067
C;Genetics:
A;Gene: SPBC20F10.11; SPDB:SPBC17D11.01
A;Map position: 2
A;Introns: 10/1; 28/2
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match             2.3%;  Score 11;  DB 2;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.0042;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 12
T42516
hypothetical protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T42516
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42516
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-427 <YOS>
A;Cross-references: EMBL:D89156; NID:g1749519; PIDN:BAA13818.1; PID:g1749520
A;Experimental source: strain PR745
C;Superfamily: hypothetical protein SPBC17D11.01

  Query Match             2.3%;  Score 11;  DB 2;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 0.0043;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308


RESULT 13
S69205
stripe a/b protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S69205
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube gui
A;Reference number: S69205; MUID:96203082; PMID:8612588
A;Accession: S69205
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1180 <FRO>
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789
C;Keywords: alternative splicing

  Query Match             2.3%;  Score 11;  DB 2;  Length 1180;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
```

```
                  ||||||||||
Db        641 HHHHHHHHHHS 651


RESULT 14
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

  Query Match             2.3%;  Score 11;  DB 1;  Length 1315;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        270 DFWGNPLRYDT 280
              ||||||||||
Db       1116 DFWGNPLRYDT 1126


RESULT 15
I39193
gene HOXA1 protein - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 15-Oct-1999
C;Accession: I39193
R;Chariot, A.; Moreau, L.; Senterre, G.; Sobel, M.; Castronovo, V.
Biochem. Biophys. Res. Commun. 215, 713-720, 1995
A;Title: Retinoic acid induces three newly cloned HOXA1 transcripts in MCF7 breast cance
A;Reference number: I39192; MUID:96011836; PMID:7488013
A;Accession: I39193
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-132 <RES>
A;Cross-references: EMBL:U37431; NID:g1051226; PIDN:AAC50250.1; PID:g1051229
C;Genetics:
A;Gene: GDB:HOXA1
A;Cross-references: GDB:120652; OMIM:142955
A;Map position: 7p15.3-7p15.3
A;Introns: 113/3
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

  Query Match             2.1%;  Score 10;  DB 2;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHH 12
              ||||||||||
Db         60 HHHHHHHHHH 69


Search completed: September 10, 2004, 05:14:57
Job time : 7.65275 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:10:39 ; Search time 4.08795 Seconds
(without alignments)
6024.823 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 473
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	447	94.5	1296	1	BXG_CLOBO	Q60393 clostridium
2	11	2.3	303	1	MOX2_HUMAN	P50222 homo sapien
3	11	2.3	303	1	MOX2_MOUSE	P32443 mus musculu
4	11	2.3	303	1	MOX2_RAT	P39020 rattus norv
5	11	2.3	354	1	OTX1_HUMAN	P32242 homo sapien
6	11	2.3	355	1	OTX1_MOUSE	P80205 mus musculu
7	11	2.3	355	1	OTX1_RAT	Q63410 rattus norv
8	11	2.3	414	1	TYY1_HUMAN	P25490 homo sapien
9	11	2.3	414	1	TYY1_MOUSE	Q00899 mus musculu
10	11	2.3	420	1	YBE1_SCHPO	O42980 schizosacch
11	11	2.3	466	1	ZIC3_MOUSE	Q62521 mus musculu
12	11	2.3	467	1	ZIC3_HUMAN	O60481 homo sapien
13	11	2.3	611	1	SNF1_CANGA	Q00372 candida gla
14	11	2.3	619	1	SNF1_CANTR	O94168 candida tro
15	11	2.3	1314	1	TETX_CLOTE	P04958 clostridium
16	10	2.1	202	1	HAN1_CHICK	Q90691 gallus gall
17	10	2.1	204	1	HAN1_SHEEP	Q28555 ovis aries
18	10	2.1	298	1	MOX2_XENLA	P39021 xenopus lae
19	10	2.1	331	1	HXA1_MOUSE	P09022 mus musculu
20	10	2.1	335	1	HXA1_HUMAN	P49639 homo sapien
21	10	2.1	351	1	HRPX_PLALO	P04929 plasmodium
22	10	2.1	375	1	CAR2_DICDI	P34907 dictyosteli
23	10	2.1	387	1	QIN_AVIS3	P56260 avian sarco
24	10	2.1	390	1	PC_DROME	P26017 drosophila
25	10	2.1	410	1	PO42_HUMAN	Q12837 homo sapien
26	10	2.1	411	1	PO42_MOUSE	Q63934 mus musculu
27	10	2.1	430	1	NOR2_RAT	Q63516 rattus norv
28	10	2.1	449	1	GAT6_HUMAN	Q92908 homo sapien
29	10	2.1	451	1	FXGB_CHICK	Q90964 gallus gall
30	10	2.1	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
31	10	2.1	465	1	HNF6_MOUSE	O08755 mus musculu
32	10	2.1	465	1	HNF6_RAT	P70512 rattus norv
33	10	2.1	469	1	FXGA_HUMAN	P55316 homo sapien
34	10	2.1	477	1	FXGB_HUMAN	P55315 homo sapien
35	10	2.1	480	1	FXGB_RAT	Q00939 rattus norv
36	10	2.1	481	1	FXGB_MOUSE	Q60987 mus musculu
37	10	2.1	551	1	CBX4_MOUSE	O55187 mus musculu
38	10	2.1	558	1	CBX4_HUMAN	O00257 homo sapien
39	10	2.1	620	1	SNF1_CANAL	P52497 candida alb
40	10	2.1	623	1	PNT1_DROME	P51022 drosophila
41	10	2.1	626	1	NR43_HUMAN	Q92570 homo sapien
42	10	2.1	628	1	NR43_RAT	P51179 rattus norv
43	10	2.1	633	1	SNF1_YEAST	P06782 saccharomyc
44	10	2.1	657	1	KNOB_PLAFN	P06719 plasmodium
45	10	2.1	665	1	LAMA_XENLA	P11048 xenopus lae

ALIGNMENTS

RESULT 1
BXG_CLOBO
ID BXG_CLOBO STANDARD; PRT; 1296 AA.
AC Q60393;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE (Bontoxilysin G).
GN BOTG.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;
RX MEDLINE=94092745; PubMed=8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins.";
RL Biochim. Biophys. Acta 1216:487-491(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X74162; CAA52275.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.

```
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             94.5%;  Score 447;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 909

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 969

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1029

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1089

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1149

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1150 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1209

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1210 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1269

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||
Db       1270 RRISENINKLRLGCNWQFIPVDEGWTE 1296


RESULT 2
MOX2_HUMAN
ID   MOX2_HUMAN     STANDARD;      PRT;   303 AA.
AC   P50222; Q9UPL6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE   homeobox).
GN   MEOX2 OR MOX2 OR GAX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95331791; PubMed=7607679;
RA   Grigoriou M., Kastrinaki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA   Pachnis V., Karagogeos D.;
RT   "Isolation of the human MOX2 homeobox gene and localization to
RT   chromosome 7p22.1-p21.3.";
RL   Genomics 26:550-555(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95229154; PubMed=7713505;
RA   Lepage D.F., Walsh K.;
RT   "Molecular cloning and localization of the human GAX gene to 7p21.";
RL   Genomics 24:535-540(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 230-303 FROM N.A.
RA   Cordes M., Lacy M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Embryo and placenta.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82629; CAA57949.1; -.
DR   EMBL; L36328; AAA58497.1; -.
DR   EMBL; BC017021; AAH17021.1; -.
DR   EMBL; AC004452; AAC06184.1; -.
DR   PIR; A55641; A55641.
DR   PIR; A56837; A56837.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04005; -.
DR   Genew; HGNC:7014; MEOX2.
DR   MIM; 600535; -.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
```

```
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DNA_BIND    186    245       HOMEOBOX.
FT   CONFLICT     58     58       G -> D (IN REF. 2).
FT   CONFLICT     79     79       MISSING (IN REF. 2).
SQ   SEQUENCE   303 AA;  33457 MW;  809ADE0CD090023D CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 3
MOX2_MOUSE
ID   MOX2_MOUSE      STANDARD;      PRT;   303 AA.
AC   P32443;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Mesenchyme homeobox 2).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201999; PubMed=1363541;
RA   Candia A.F., Hu J., Crosby J., Lalley P.A., Noden D., Nadeau J.H.,
RA   Wright C.V.E.;
RT   "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are
RT   differentially expressed during early mesodermal patterning in mouse
RT   embryos.";
RL   Development 116:1123-1136(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94232829; PubMed=7909944;
RA   Candia A.F., Kovalik J.-P., Wright C.V.E.;
RT   "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat
RT   homologs.";
RL   Nucleic Acids Res. 21:4982-4982(1993).
RN   [3]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=95349593; PubMed=7623821;
RA   Andres V., Fisher S., Wearsch P., Walsh K.;
RT   "Regulation of Gax homeobox gene transcription by a combination of
RT   positive factors including myocyte-specific enhancer factor 2.";
RL   Mol. Cell. Biol. 15:4272-4281(1995).
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-
CC       8.5 dpc it is found on the entire epithelium of the somite. At 9.5
CC       dpc its expression is restricted to the sclerotome. At 10.5 dpc it
CC       is found in sclerotomally derived cells including the vertebral
CC       and costal precursors.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16406; CAA78899.1; -.
DR   EMBL; S79168; -; NOT_ANNOTATED_CDS.
DR   PIR; B49122; B49122.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04048; -.
DR   MGD; MGI:103219; Meox2.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       63     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33506 MW;  41BD05FC39AA4427 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 4
MOX2_RAT
ID   MOX2_RAT        STANDARD;      PRT;   303 AA.
AC   P39020;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein MOX-2 (Growth arrest-specific homeobox).
GN   MEOX2 OR MOX2 OR MOX-2 OR GAX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=93268321; PubMed=8098844;
RA   Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,
RA   Walsh K.;
RT   "Molecular cloning of a diverged homeobox gene that is rapidly down-
RT   regulated during the G0/G1 transition in vascular smooth muscle
RT   cells.";
RL   Mol. Cell. Biol. 13:3722-3733(1993).
RN   [2]
RP   REVISIONS.
RA   Walsh K.;
RL   Submitted (MAR-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Role in mesoderm induction and its earliest regional
CC       specification, somitogenesis, and myogenic and sclerotomal
CC       differentiation. May have a regulatory role when quiescent
CC       vascular smooth muscle cells reenter the cell cycle.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Aorta and heart. Also detected in lung and
CC       kidney.
CC   -!- INDUCTION: Rapidly and transiently down-regulated during the
CC       transition from G0 to G1 induced by mitogen stimulation.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z17223; CAA78931.1; -.
DR   PIR; A48130; A48130.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T04054; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN       42     47       POLY-SER.
FT   DOMAIN       68     79       POLY-HIS.
FT   DOMAIN       80     85       POLY-GLN.
FT   DOMAIN       64     85       GLN/HIS-RICH (OPA-REPEAT).
FT   DNA_BIND    186    245       HOMEOBOX.
SQ   SEQUENCE   303 AA;  33605 MW;  7776642AEFA3A2E8 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 5
OTX1_HUMAN
ID   OTX1_HUMAN     STANDARD;      PRT;   354 AA.
AC   P32242;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS):
CC       5'-TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; BC007621; AAH07621.1; -.
DR   PIR; S39406; S39406.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02079; -.
DR   Genew; HGNC:8521; OTX1.
DR   MIM; 600036; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007275; P:development; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    301       HIS-RICH.
SQ   SEQUENCE   354 AA;  37327 MW;  E32C1E47464BDDAA CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 0.0069;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 6
OTX1_MOUSE
ID   OTX1_MOUSE     STANDARD;      PRT;   355 AA.
AC   P80205;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1 OR OTX-1.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=93327763; PubMed=8101484;
RA   Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M.R.,
RA   Nigro V., Boncinelli E.;
RT   "A vertebrate gene related to orthodenticle contains a homeodomain of
RT   the bicoid class and demarcates anterior neuroectoderm in the
RT   gastrulating mouse embryo.";
RL   EMBO J. 12:2735-2747(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 31-98 FROM N.A.
RX   MEDLINE=92357147; PubMed=1353865;
RA   Simeone A., Acampora D., Gulisano M., Stornaiuolo A., Boncinelli E.;
RT   "Nested expression domains of four homeobox genes in developing
RT   rostral brain.";
RL   Nature 358:687-690(1992).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve.
CC   -!- DEVELOPMENTAL STAGE: Embryo.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC057105; AAH57105.1; -.
DR   EMBL; BC058354; AAH58354.1; -.
DR   EMBL; X68883; CAA48754.1; -.
DR   PIR; S35345; S35345.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T02080; -.
DR   MGD; MGI:97450; Otx1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
SQ   SEQUENCE   355 AA;  37531 MW;  DA5136D9D4FC948F CRC64;

  Query Match              2.3%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0069;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 7
OTX1_RAT
ID   OTX1_RAT       STANDARD;      PRT;   355 AA.
AC   Q63410; Q64203;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Homeobox protein OTX1.
GN   OTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95016961; PubMed=7931541;
RA   Frantz G.D., Weimann J.M., Levin M.E., McConnell S.K.;
RT   "Otx1 and Otx2 define layers and regions in developing cerebral
RT   cortex and cerebellum.";
RL   J. Neurosci. 14:5725-5740(1994).
RN   [2]
RP   SEQUENCE OF 176-323 FROM N.A.
RX   MEDLINE=96108898; PubMed=8613727;
RA   Robel L., Ding M., James A.J., Lin X., Simeone A., Leckman J.F.,
RA   Vaccarino F.M.;
RT   "Fibroblast growth factor 2 increases Otx2 expression in precursor
RT   cells from mammalian telencephalon.";
RL   J. Neurosci. 15:7879-7891(1995).
CC   -!- FUNCTION: Probably play a role in the development of the brain and
CC       the sense organs. Can bind to the BCD target sequence (BTS): 5'-
CC       TCTAATCCC-3' (By similarity). May play a role in the specification
CC       or differentiation of neurons in the deep layers of the cerebral
CC       cortex, and also in cerebellar regionalization during early
CC       development.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Brain: restricted regions of the developing
CC       rostral brain including the presumptive cerebral cortex and
CC       olfactory bulbs; expressed in the developing olfactory, auricolar
CC       and ocular systems, including the covering of the optic nerve (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: Expressed throughout the forebrain and
CC       midbrain during development, and in addition is also seen in
CC       discrete spatial and temporal domains in the developing cerebral
CC       cortex and cerebellum. Confined to a subpopulation of neurons in
CC       layers 5 and 6 within the adult cerebral cortex and during
CC       development expression is high in the progenitors of these deep-
CC       layer cells. Expressed in the developing cerebellum in spatially
CC       restricted regions of the external granular layer.
CC   -!- SIMILARITY: Belongs to the paired homeobox family. Bicoid
```

```
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32602; AAA53557.1; -.
DR   EMBL; S81924; -; NOT_ANNOTATED_CDS.
DR   PIR; I56547; I56547.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR003025; Otx_TF.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     38     97       HOMEOBOX.
FT   DOMAIN      275    302       HIS-RICH.
FT   CONFLICT    219    219       T -> A (IN REF. 2).
FT   CONFLICT    229    229       R -> G (IN REF. 2).
SQ   SEQUENCE   355 AA;  37602 MW;  C875871723D0B876 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0.0069;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        291 GHHHHHHHHHH 301


RESULT 8
TYY1_HUMAN
ID   TYY1_HUMAN     STANDARD;      PRT;   414 AA.
AC   P25490; Q14935;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1).
GN   YY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92005716; PubMed=1655281;
RA   Shi Y., Seto E., Chang L.-S., Shenk T.;
RT   "Transcriptional repression by YY1, a human GLI-Kruppel-related
RT   protein, and relief of repression by adenovirus E1A protein.";
RL   Cell 67:377-388(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=92052179; PubMed=1946405;
RA   Park K., Atchison M.;
RT   "Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta),
RT   that binds to the immunoglobulin kappa 3' enhancer and the
RT   immunoglobulin heavy-chain mu E1 site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9804-9808(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Whitson R.H., Huang T., Dang J., Itakura K.;
RL   Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=98152931; PubMed=9493912;
RA   McNeil S., Guo B., Stein J.L., Lian J.B., Bushmeyer S., Seto E.,
RA   Atchison M.L., Penman S., van Wijnen A.J., Stein G.S.;
RT   "Targeting of the YY1 transcription factor to the nucleolus and the
RT   nuclear matrix in situ: the C-terminus is a principal determinant for
RT   nuclear trafficking.";
RL   J. Cell. Biochem. 68:500-510(1998).
RN   [6]
RP   INTERACTION WITH YAF2.
RX   MEDLINE=97169297; PubMed=9016636;
RA   Kalenik J.L., Chen D., Bradley M.E., Chen S.-J., Lee T.-C.;
RT   "Yeast two-hybrid cloning of a novel zinc finger protein that
RT   interacts with the multifunctional transcription factor YY1.";
RL   Nucleic Acids Res. 25:843-849(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 304-414.
RX   MEDLINE=97098436; PubMed=8942976;
RA   Houbaviy H.B., Usheva A., Shenk T., Burley S.K.;
RT   "Cocrystal structure of YY1 bound to the adeno-associated virus P5
RT   initiator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13577-13582(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 353-379.
RX   MEDLINE=98308000; PubMed=9642075;
RA   Viles J.H., Patel S.U., Mitchell J.B.O., Moody C.M., Justice D.E.,
RA   Uppenbrink J., Doyle P.M., Harris C.J., Sadler P.J., Thornton J.M.;
RT   "Design, synthesis and structure of a zinc finger with an artificial
RT   beta-turn.";
RL   J. Mol. Biol. 279:973-986(1998).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. FOR
CC       EXAMPLE IT ACTS AS A REPRESSOR IN ABSENCE OF ADENOVIRUS E1A
CC       PROTEIN BUT AS AN ACTIVATOR IN ITS PRESENCE.
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
CC       first and second zinc fingers.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M77698; AAA59467.1; -.
DR   EMBL; M76541; AAA59926.1; -.
DR   EMBL; Z14077; CAA78455.1; -.
DR   EMBL; BC037308; AAH37308.1; -.
DR   PIR; A40350; A40350.
DR   PDB; 1UBD; 23-DEC-96.
DR   PDB; 1ZNM; 01-APR-98.
DR   TRANSFAC; T00915; -.
DR   Genew; HGNC:12856; YY1.
DR   MIM; 600013; -.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; 3D-structure.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     69       GLY-RICH.
FT   DOMAIN       70     80       POLY-HIS.
FT   DOMAIN      159    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE 1.
FT   ZN_FING     325    347       C2H2-TYPE 2.
FT   ZN_FING     353    377       C2H2-TYPE 3.
FT   ZN_FING     383    407       C2H2-TYPE 4.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION.
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION.
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN.
FT   CONFLICT     65     65       H -> R (IN REF. 2).
FT   CONFLICT    196    196       G -> R (IN REF. 1).
FT   STRAND      296    297
FT   TURN        301    302
FT   STRAND      306    307
FT   HELIX       310    317
FT   HELIX       318    320
FT   STRAND      325    326
FT   TURN        328    330
FT   STRAND      333    334
FT   HELIX       337    343
FT   HELIX       344    346
FT   TURN        347    347
FT   STRAND      353    354
FT   TURN        358    359
FT   STRAND      363    364
FT   HELIX       367    378
FT   TURN        388    389
FT   TURN        397    400
FT   HELIX       401    407
SQ   SEQUENCE   414 AA;  44712 MW;  058C05A0AD2D04E6 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0079;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         69 GHHHHHHHHHH 79
```

```
RESULT 9
TYY1_MOUSE
ID   TYY1_MOUSE     STANDARD;      PRT;   414 AA.
AC   Q00899;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional repressor protein YY1 (Yin and yang 1) (YY-1) (Delta
DE   transcription factor) (NF-E1) (UCR-motif DNA-binding protein).
GN   YY1 OR UCRBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107191; PubMed=1309593;
RA   Flanagan J.R., Becker K.G., Ennist D.L., Gleason S.L., Driggers P.H.,
RA   Levi B.-Z., Appella E., Ozato K.;
RT   "Cloning of a negative transcription factor that binds to the
RT   upstream conserved region of Moloney murine leukemia virus.";
RL   Mol. Cell. Biol. 12:38-44(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93296177; PubMed=8516301;
RA   Safrany G., Perry R.P.;
RT   "Characterization of the mouse gene that encodes the delta/YY1/NF-
RT   E1/UCRBP transcription factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:5559-5563(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052178; PubMed=1946404;
RA   Hariharan N., Kelley D.E., Perry R.P.;
RT   "Delta, a transcription factor that binds to downstream elements in
RT   several polymerase II promoters, is a functionally versatile zinc
RT   finger protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9799-9803(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: MULTIFUNCTIONAL TRANSCRIPTION FACTOR THAT EXHIBITS
CC       POSITIVE AND NEGATIVE CONTROL ON A LARGE NUMBER OF CELLULAR AND
CC       VIRAL GENES BY BINDING TO SITES OVERLAPPING THE TRANSCRIPTION
CC       START SITE. MAY PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC       DIFFERENTIATION. THE FUNCTION OF YY1 AS AN ACTIVATOR OR A
CC       REPRESSOR IS SPECIFIED BY THE PRESENCE OF OTHER PROTEINS. BINDS TO
CC       THE UPSTREAM CONSERVED REGION (UCR) (5'-CGCCATTTT-3') OF MOLONEY
CC       MURINE LEUKEMIA VIRUS (MULV).
CC   -!- SUBUNIT: Interacts with YAF2 through the region encompassing the
```

```
CC       first and second zinc fingers (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M73963; AAA40522.1; -.
DR   EMBL; L13968; AAA40477.1; -.
DR   EMBL; L13969; AAA40477.1; JOINED.
DR   EMBL; L13965; AAA40477.1; JOINED.
DR   EMBL; L13966; AAA40477.1; JOINED.
DR   EMBL; L13967; AAA40477.1; JOINED.
DR   EMBL; M74590; AAA37521.1; -.
DR   EMBL; BC055899; AAH55899.1; -.
DR   PIR; A48273; A48273.
DR   HSSP; P25490; 1UBD.
DR   TRANSFAC; T00278; -.
DR   TRANSFAC; T00865; -.
DR   MGD; MGI:99150; Yy1.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003700; F:transcription factor activity; IDA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Repressor; Activator; Nuclear protein;
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   DOMAIN       43     53       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       54     70       GLY-RICH.
FT   DOMAIN       71     82       POLY-HIS.
FT   DOMAIN      161    170       GLY/SER-RICH.
FT   ZN_FING     296    320       C2H2-TYPE.
FT   ZN_FING     325    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    407       C2H2-TYPE.
FT   DOMAIN      257    341       INVOLVED IN NUCLEAR MATRIX ASSOCIATION
FT                                (BY SIMILARITY).
FT   DOMAIN      333    371       INVOLVED IN REPRESSION OF ACTIVATED
FT                                TRANSCRIPTION (BY SIMILARITY).
FT   DOMAIN      371    397       INVOLVED IN MASKING TRANSACTIVATION
FT                                DOMAIN (BY SIMILARITY).
FT   CONFLICT    219    219       F -> S (IN REF. 3).
FT   CONFLICT    375    375       R -> G (IN REF. 3).
SQ   SEQUENCE   414 AA;  44717 MW;  C012378288E984F9 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 0.0079;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         70 GHHHHHHHHHH 80


RESULT 10
YBE1_SCHPO
ID   YBE1_SCHPO     STANDARD;      PRT;   420 AA.
AC   O42980; P78807;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative thiol protease C17D11.01 (EC 3.4.22.-).
GN   SPBC17D11.01 OR SPBC20F10.11.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.;
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- SIMILARITY: Belongs to peptidase family C48.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL031739; CAA21072.1; -.
DR   EMBL; AL021747; CAA16851.1; -.
DR   EMBL; D89156; BAA13818.1; -.
DR   GeneDB_SPombe; SPBC17D11.01; -.
DR   InterPro; IPR003653; Peptidase_C48.
DR   Pfam; PF02902; Peptidase_C48; 1.
DR   PROSITE; PS50600; ULP_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Thiol protease.
FT   DOMAIN      299    312       POLY-HIS.
FT   CONFLICT     57     57       L -> W (IN REF. 2).
FT   CONFLICT    420    420       F -> FGCYTCVV (IN REF. 2).
SQ   SEQUENCE   420 AA;  47257 MW;  2125D94D9A2D9E87 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 420;
  Best Local Similarity   100.0%;  Pred. No. 0.008;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        298 GHHHHHHHHHH 308
```

```
RESULT 11
ZIC3_MOUSE
ID   ZIC3_MOUSE     STANDARD;      PRT;   466 AA.
AC   Q62521;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=96132843; PubMed=8557628;
RA   Aruga J., Nagai T., Tokuyama T., Hayashizaki Y., Okazaki Y.,
RA   Chapman V.M., Mikoshiba K.;
RT   "The mouse zic gene family. Homologues of the Drosophila pair-rule
RT   gene odd-paired.";
RL   J. Biol. Chem. 271:1043-1047(1996).
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: CNS. A high level expression is seen in the
CC       cerebellum.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D70849; BAA11116.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T04671; -.
DR   MGD; MGI:106676; Zic3.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     96       POLY-HIS.
FT   ZN_FING     294    321       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     327    351       C2H2-TYPE 2.
FT   ZN_FING     357    381       C2H2-TYPE 3.
FT   ZN_FING     387    409       C2H2-TYPE 4.
SQ   SEQUENCE   466 AA;  50644 MW;  089749DF726F069B CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 0.0088;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 12
```

```
ZIC3_HUMAN
ID   ZIC3_HUMAN     STANDARD;      PRT;   467 AA.
AC   O60481;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein ZIC 3 (Zinc finger protein of the cerebellum 3).
GN   ZIC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HTX1 MET-323.
RX   MEDLINE=98016414; PubMed=9354794;
RA   Gebbia M., Ferrero G.B., Pilia G., Bassi M.T., Aylsworth A.S.,
RA   Penman-Splitt M., Bird L.M., Bamforth J.S., Burn J., Schlessiner D.,
RA   Nelson D.L., Casey B.;
RT   "X-linked situs abnormalities result from mutations in ZIC3.";
RL   Nat. Genet. 17:305-308(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably functions as a transcription factor in the
CC       earliest stages of the left-right (LR) body axis formation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in ZIC3 are the cause of X-Linked visceral
CC       heterotaxy (HTX1) [MIM:306955]. HTX1 is a disease characterized by
CC       congenital heart disease and alterations of visceral situs,
CC       including asplenia or polysplenia, symmetric liver, intestinal
CC       malrotation, and abnormal lung lobation.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF028706; AAC05594.1; -.
DR   EMBL; AL035443; CAB41648.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:12874; ZIC3.
DR   MIM; 300265; -.
DR   MIM; 306955; -.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 4.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   Repeat; Nuclear protein; Disease mutation.
FT   DOMAIN       46     55       POLY-ALA.
FT   DOMAIN       87     97       POLY-HIS.
FT   ZN_FING     295    322       C2H2-TYPE 1 (ATYPICAL).
FT   ZN_FING     328    352       C2H2-TYPE 2.
FT   ZN_FING     358    382       C2H2-TYPE 3.
FT   ZN_FING     388    410       C2H2-TYPE 4.
FT   VARIANT     323    323       T -> M (in HTX1).
FT                                /FTId=VAR_007753.
SQ   SEQUENCE   467 AA;  50569 MW;  3150CF13C0679568 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 0.0088;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         86 GHHHHHHHHHH 96


RESULT 13
SNF1_CANGA
ID   SNF1_CANGA     STANDARD;      PRT;   611 AA.
AC   Q00372;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida glabrata (Yeast) (Torulopsis glabrata).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5478;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCCLS84;
RX   MEDLINE=97101049; PubMed=8945576;
RA   Petter R., Kwon-Chung K.J.;
RT   "Disruption of the SNF1 gene abolishes trehalose utilization in the
RT   pathogenic yeast Candida glabrata.";
RL   Infect. Immun. 64:5269-5273(1996).
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78130; AAB48642.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN        6     17       POLY-HIS.
FT   DOMAIN       39    290       PROTEIN KINASE.
FT   NP_BIND      45     53       ATP (BY SIMILARITY).
FT   BINDING      68     68       ATP (BY SIMILARITY).
FT   ACT_SITE    161    161       BY SIMILARITY.
FT   MOD_RES     194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  70049 MW;  89E17812A4900CD0 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db          8 HHHHHHHHHHS 18


RESULT 14
SNF1_CANTR
ID   SNF1_CANTR     STANDARD;      PRT;   619 AA.
AC   O94168;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN   SNF1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanai T., Ogawa K., Ueda M., Tanaka A.;
RT   "Genetic evaluation of the function of SNF1 in Candida tropicalis.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential for release from glucose repression. It
CC       interacts and has functional relationship to the regulatory
CC       protein SNF4. Could phosphorylates CAT8 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Associated with the nuclear membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. SNF1
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB024535; BAA75889.1; -.
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT   DOMAIN        17     29       POLY-HIS.
FT   DOMAIN        52    303       PROTEIN KINASE.
FT   NP_BIND       58     66       ATP (BY SIMILARITY).
FT   BINDING       81     81       ATP (BY SIMILARITY).
FT   ACT_SITE     174    174       BY SIMILARITY.
FT   MOD_RES      207    207       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   619 AA;  70323 MW;  0FCF1FC3DCE706D7 CRC64;

  Query Match             2.3%;  Score 11;  DB 1;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHS 13
              ||||||||||
Db         20 HHHHHHHHHHS 30


RESULT 15
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE   chain (Tetanus toxin chain H)].
```

```
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
FT   TURN        928    929
```

FT	STRAND	932	935
FT	HELIX	938	940
FT	TURN	941	946
FT	STRAND	949	956
FT	HELIX	962	968
FT	TURN	969	970
FT	STRAND	972	977
FT	STRAND	980	981
FT	HELIX	983	985
FT	STRAND	987	995
FT	TURN	996	997
FT	STRAND	998	1004
FT	TURN	1006	1007
FT	STRAND	1010	1016
FT	STRAND	1020	1020
FT	TURN	1021	1022
FT	STRAND	1031	1037
FT	TURN	1039	1040
FT	STRAND	1042	1047
FT	TURN	1048	1049
FT	STRAND	1050	1056
FT	TURN	1058	1059
FT	STRAND	1068	1074
FT	TURN	1079	1080
FT	STRAND	1082	1091
FT	HELIX	1097	1105
FT	TURN	1106	1107
FT	STRAND	1112	1112
FT	STRAND	1114	1114
FT	TURN	1116	1117
FT	STRAND	1120	1120
FT	STRAND	1122	1122
FT	TURN	1123	1124
FT	STRAND	1127	1131
FT	HELIX	1132	1134
FT	TURN	1135	1136
FT	STRAND	1137	1141
FT	TURN	1144	1145
FT	STRAND	1148	1152
FT	STRAND	1155	1158
FT	TURN	1159	1162
FT	STRAND	1163	1166
FT	STRAND	1173	1178
FT	TURN	1184	1185
FT	STRAND	1188	1188
FT	STRAND	1190	1190

Query Match 2.3%; Score 11; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        270 DFWGNPLRYDT 280
              ||||||||||| 
Db       1115 DFWGNPLRYDT 1125
```

Search completed: September 10, 2004, 05:07:25
Job time : 5.08795 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 04:18:27 ; Search time 19.2374 Seconds
(without alignments)
7757.804 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 473
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	12	2.5	226	11	Q9QZD2	Q9qzd2 rattus norv
2	12	2.5	516	13	O42186	O42186 brachydanio
3	12	2.5	517	13	Q90YL6	Q90yl6 brachydanio
4	12	2.5	569	13	Q802D2	Q802d2 brachydanio
5	12	2.5	679	13	Q803K9	Q803k9 brachydanio
6	12	2.5	706	5	Q8SXG1	Q8sxg1 drosophila
7	12	2.5	861	5	Q9VRT8	Q9vrt8 drosophila
8	11	2.3	163	5	Q7YTA3	Q7yta3 glomeris ma
9	11	2.3	225	13	Q90XZ3	Q90xz3 erpetoichth
10	11	2.3	230	4	O75263	O75263 homo sapien
11	11	2.3	243	4	Q9P2R2	Q9p2r2 homo sapien
12	11	2.3	243	6	Q9N2G2	Q9n2g2 gorilla gor
13	11	2.3	243	6	Q9N2G3	Q9n2g3 pan troglod
14	11	2.3	243	6	Q9N2G1	Q9n2g1 pongo pygma
15	11	2.3	302	13	Q90YH7	Q90yh7 gallus gall
16	11	2.3	303	6	Q95JA6	Q95ja6 sus scrofa
17	11	2.3	303	11	Q99M23	Q99m23 mus musculu
18	11	2.3	355	11	Q91ZJ3	Q91zj3 mus musculu
19	11	2.3	358	13	Q9DG32	Q9dg32 xenopus lae
20	11	2.3	411	11	Q8CHJ4	Q8chj4 rattus norv
21	11	2.3	441	5	Q8T1W3	Q8t1w3 dictyosteli
22	11	2.3	449	10	Q8L4M9	Q8l4m9 oryza sativ
23	11	2.3	451	2	Q9LA13	Q9la13 clostridium
24	11	2.3	456	11	Q8BSB3	Q8bsb3 mus musculu
25	11	2.3	467	13	Q9YGC6	Q9ygc6 xenopus lae
26	11	2.3	511	10	Q8GXZ5	Q8gxz5 arabidopsis
27	11	2.3	906	5	Q8IN91	Q8in91 drosophila
28	11	2.3	1122	5	Q8SXD4	Q8sxd4 drosophila
29	11	2.3	1147	5	Q95PH8	Q95ph8 dictyosteli
30	11	2.3	1180	5	Q24163	Q24163 drosophila
31	11	2.3	1186	5	Q9VED3	Q9ved3 drosophila
32	11	2.3	1283	5	Q95RH4	Q95rh4 drosophila
33	11	2.3	1310	2	Q93N27	Q93n27 clostridium
34	11	2.3	1911	5	Q9W4M7	Q9w4m7 drosophila
35	10	2.1	53	2	O55244	O55244 escherichia
36	10	2.1	233	4	Q8TC53	Q8tc53 homo sapien
37	10	2.1	233	5	Q8MP30	Q8mp30 dictyosteli
38	10	2.1	236	11	Q9QWE1	Q9qwe1 mus sp. cla
39	10	2.1	239	5	Q7Z1G6	Q7z1g6 trypanosoma
40	10	2.1	244	13	Q8JHA3	Q8jha3 gallus gall
41	10	2.1	254	5	Q9V582	Q9v582 drosophila
42	10	2.1	298	13	Q7ZXL2	Q7zxl2 xenopus lae
43	10	2.1	309	5	Q18751	Q18751 caenorhabdi
44	10	2.1	318	10	Q8H3Q1	Q8h3q1 oryza sativ
45	10	2.1	321	10	O23891	O23891 oryza sativ

ALIGNMENTS

RESULT 1
Q9QZD2
ID Q9QZD2 PRELIMINARY; PRT; 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Collagen XVIII (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20227226; PubMed=10766159;
RA Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA Folkman J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat
RT primary mammary tumors.";
RL Cancer Res. 60:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -.
DR HSSP; P39061; 1KOE.
FT NON_TER 1 1
SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0E949 CRC64;

Query Match 2.5%; Score 12; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.00095;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db         17 HHHHHHHHHHSS 28
```

RESULT 2
O42186
ID O42186 PRELIMINARY; PRT; 516 AA.
AC O42186;

```
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FTZ-F1 homolog.
GN   NR5A2 OR ZFF1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322094; PubMed=9178748;
RA   Liu D., Le Drean Y., Ekker M., Xiong F., Hew C.L.;
RT   "Teleost FTZ-F1 homolog and its splicing variant determine the
RT   expression of the salmon gonadotropin IIbeta subunit gene.";
RL   Mol. Endocrinol. 11:877-890(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AF014926; AAC60274.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04065; -.
DR   ZFIN; ZDB-GENE-990415-79; nr5a2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   516 AA;  58805 MW;  500DA09E116D5FA5 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 3
Q90YL6
ID   Q90YL6      PRELIMINARY;      PRT;   517 AA.
AC   Q90YL6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nuclear receptor FF1aE.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chai C., Tan J.H., Gao W., Teh H.L., Chan W.K.;
RT   "Full length zebrafish ff1aE cDNA.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
DR   EMBL; AY034061; AAK54449.1; -.
DR   HSSP; P19793; 1LBD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003707; F:steroid hormone receptor activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Metal-binding; Nuclear protein; Receptor; Transcription;
KW   Transcription regulation; Zinc; Zinc-finger.
SQ   SEQUENCE   517 AA;  58761 MW;  CF48D08AB7A75F29 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        189 HHHHHHHHHHSS 200


RESULT 4
Q802D2
ID   Q802D2      PRELIMINARY;      PRT;   569 AA.
AC   Q802D2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048886; AAH48886.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
KW   Cyclin.
SQ   SEQUENCE   569 AA;  64117 MW;  812A7332C06EF4C2 CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        486 HHHHHHHHHHSS 497


RESULT 5
Q803K9
ID   Q803K9      PRELIMINARY;      PRT;   679 AA.
AC   Q803K9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cyclin T1 (Fragment).
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044435; AAH44435.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 1.
KW   Cyclin.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  74887 MW;  AB5667D59439659F CRC64;

  Query Match             2.5%;  Score 12;  DB 13;  Length 679;
  Best Local Similarity   100.0%;  Pred. No. 0.0026;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HHHHHHHHHHSS 14
              ||||||||||||
Db        488 HHHHHHHHHHSS 499


RESULT 6
Q8SXG1
ID   Q8SXG1      PRELIMINARY;      PRT;   706 AA.
AC   Q8SXG1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RH12258p.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AY089654; AAL90392.1; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   706 AA;  79683 MW;  73C0F8199DF60E5F CRC64;

  Query Match             2.5%;  Score 12;  DB 5;  Length 706;
  Best Local Similarity   100.0%;  Pred. No. 0.0026;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        370 GHHHHHHHHHHS 381


RESULT 7
Q9VRT8
ID   Q9VRT8      PRELIMINARY;      PRT;   861 AA.
AC   Q9VRT8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG6619 protein.
GN   CG6619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AE003564; AAF50702.2; -.
DR   FlyBase; FBgn0035676; CG6619.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   861 AA;  96809 MW;  8530ED5D92671A41 CRC64;

  Query Match             2.5%;  Score 12;  DB 5;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 0.0032;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHHS 13
              ||||||||||||
Db        525 GHHHHHHHHHHS 536
```

```
RESULT 8
Q7YTA3
ID   Q7YTA3      PRELIMINARY;      PRT;   163 AA.
AC   Q7YTA3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Distal-less protein (Fragment).
GN   DLL.
OS   Glomeris marginata.
OC   Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Pentazonia;
OC   Glomerida; Glomeridae; Glomeris.
OX   NCBI_TaxID=62006;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22767835; PubMed=12885558;
RA   Prpic N.M., Tautz D.;
RT   "The expression of the proximodistal axis patterning genes Distal-less
RT   and dach shund in the appendages of Glomeris marginata (Myriapoda:
RT   Diplopoda) suggests a special role for these genes in patterning the
RT   head appendages.";
RL   Dev. Biol. 260:97-112(2003).
DR   EMBL; AJ551276; CAD82905.1; -.
FT   NON_TER       1      1
FT   NON_TER     163    163
SQ   SEQUENCE   163 AA;  17588 MW;  87B8BCD9F5421697 CRC64;

  Query Match             2.3%;  Score 11;  DB 5;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 0.0079;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHH 11
              ||||||||||
Db         16 MGHHHHHHHHH 26


RESULT 9
Q90XZ3
ID   Q90XZ3      PRELIMINARY;      PRT;   225 AA.
AC   Q90XZ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Homeodomain protein Otx1 (Fragment).
OS   Erpetoichthys calabaricus (Rope fish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Polypteriformes; Polypteridae; Erpetoichthys.
OX   NCBI_TaxID=27687;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21396394; PubMed=11504847;
RA   Germot A., Lecointre G., Plouhinec J.L., Le Mentec C., Girardot F.,
RA   Mazan S.;
RT   "Structural evolution of otx genes in Craniates.";
RL   Mol. Biol. Evol. 18:1668-1678(2001).
DR   EMBL; AF321406; AAL01673.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     225    225
SQ   SEQUENCE   225 AA;  23475 MW;  0AE78B8DB48B4941 CRC64;

  Query Match             2.3%;  Score 11;  DB 13;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        172 GHHHHHHHHHH 182


RESULT 10
O75263
ID   O75263      PRELIMINARY;      PRT;   230 AA.
AC   O75263;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   WUGSC:H_DJ0620P06.1 protein (Fragment).
GN   WUGSC:H_DJ0620P06.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Andrews S., Lacy M., Dubbelde C.;
RT   "The sequence of Homo sapiens PAC clone RP4-620P6.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005550; AAC33152.1; -.
DR   HSSP; P14653; 1B72.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
FT   NON_TER     230    230
SQ   SEQUENCE   230 AA;  25476 MW;  81D627D0CFEC00B0 CRC64;

  Query Match             2.3%;  Score 11;  DB 4;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


RESULT 11
Q9P2R2
ID   Q9P2R2      PRELIMINARY;      PRT;   243 AA.
AC   Q9P2R2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Homeobox protein OTX1 (Fragment).
GN   OTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037501; BAA90421.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
```

```
FT   NON_TER     243    243
SQ   SEQUENCE   243 AA;  24428 MW;  01FA60D616BBAF65 CRC64;

  Query Match             2.3%;  Score 11;  DB 4;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        190 GHHHHHHHHHH 200


RESULT 12
Q9N2G2
ID   Q9N2G2      PRELIMINARY;      PRT;   243 AA.
AC   Q9N2G2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Homeobox protein OTX1 (Fragment).
GN   OTX1.
OS   Gorilla gorilla (gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9593;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=gorilla-U1;
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037503; BAA90423.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     243    243
SQ   SEQUENCE   243 AA;  24428 MW;  01FA60D616BBAF65 CRC64;

  Query Match             2.3%;  Score 11;  DB 6;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        190 GHHHHHHHHHH 200


RESULT 13
Q9N2G3
ID   Q9N2G3      PRELIMINARY;      PRT;   243 AA.
AC   Q9N2G3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Homeobox protein OTX1 (Fragment).
GN   OTX1.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=chimp-220;
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037502; BAA90422.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     243    243
SQ   SEQUENCE   243 AA;  24428 MW;  01FA60D616BBAF65 CRC64;

  Query Match             2.3%;  Score 11;  DB 6;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        190 GHHHHHHHHHH 200


RESULT 14
Q9N2G1
ID   Q9N2G1      PRELIMINARY;      PRT;   243 AA.
AC   Q9N2G1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Homeobox protein OTX1 (Fragment).
GN   OTX1.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=oran-U1;
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037504; BAA90424.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR003025; Otx_TF.
DR   Pfam; PF03529; TF_Otx; 1.
DR   PRINTS; PR01255; OTXHOMEOBOX.
FT   NON_TER       1      1
FT   NON_TER     243    243
SQ   SEQUENCE   243 AA;  24398 MW;  004A60CD1700AFC5 CRC64;

  Query Match             2.3%;  Score 11;  DB 6;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db        190 GHHHHHHHHHH 200


RESULT 15
Q90YH7
ID   Q90YH7      PRELIMINARY;      PRT;   302 AA.
AC   Q90YH7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   MOX-2 protein.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21297712; PubMed=11404088;
RA   Rallis C., Stamataki D., Pontikakis S., Mankoo B., Karagogeos D.;
```

```
RT   "Isolation of the avian homologue of the homeobox gene Mox2 and
RT   analysis of its expression pattern in developing somites and limbs.";
RL   Mech. Dev. 104:121-124(2001).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AJ401088; CAC42831.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
SQ   SEQUENCE   302 AA;  33414 MW;  C359043CF6B5CD7A CRC64;

  Query Match             2.3%;  Score 11;  DB 13;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 0.014;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GHHHHHHHHHH 12
              ||||||||||
Db         67 GHHHHHHHHHH 77


Search completed: September 10, 2004, 05:12:59
Job time : 19.2374 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 02:14:55 ; Search time 31.2007 Seconds
(without alignments)
4283.398 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 2550
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2550	100.0	473	2	AAW68400	Aaw68400 Clostridi
2	2391	93.8	449	3	AAY77139	Aay77139 Synthetic
3	2391	93.8	449	4	AAB04167	Aab04167 Botulism
4	1369.5	53.7	472	2	AAW68394	Aaw68394 Clostridi
5	1350.5	53.0	472	2	AAW68393	Aaw68393 Clostridi
6	1211.5	47.5	1291	2	AAW68392	Aaw68392 Clostridi
7	1206.5	47.3	1070	3	AAY93308	Aay93308 A mangane
8	1206.5	47.3	1095	3	AAY93311	Aay93311 A mangane
9	1175.5	46.1	439	4	AAB04085	Aab04085 Botulism
10	1168.5	45.8	440	3	AAY77135	Aay77135 Synthetic
11	1168.5	45.8	440	4	AAB04091	Aab04091 Botulism
12	1143.5	44.8	848	4	AAB04082	Aab04082 Botulism
13	965	37.8	462	2	AAR95009	Aar95009 Type A ne
14	965	37.8	462	2	AAW68390	Aaw68390 Clostridi
15	900	35.3	452	2	AAW68396	Aaw68396 Clostridi
16	894	35.1	1169	2	AAW56017	Aaw56017 Recombina
17	885.5	34.7	451	2	AAW68395	Aaw68395 Clostridi
18	844.5	33.1	445	2	AAW68391	Aaw68391 Clostridi
19	815.5	32.0	448	2	AAW68399	Aaw68399 Clostridi
20	815	32.0	437	4	AAB04088	Aab04088 Botulism
21	815	32.0	438	2	AAR95008	Aar95008 Type A ne
22	815	32.0	438	2	AAW68389	Aaw68389 Clostridi
23	815	32.0	438	3	AAY77134	Aay77134 Synthetic
24	815	32.0	1067	3	AAY93307	Aay93307 A mangane
25	815	32.0	1092	3	AAY93310	Aay93310 A mangane
26	815	32.0	1296	2	AAR95010	Aar95010 C. botuli
27	810.5	31.8	435	4	AAB04090	Aab04090 Botulism
28	810	31.8	434	4	AAB04089	Aab04089 Botulism
29	801	31.4	432	3	AAY77142	Aay77142 Native bo
30	801	31.4	837	3	AAY77140	Aay77140 Native bo
31	800	31.4	847	4	AAB04081	Aab04081 Botulism
32	798	31.3	1295	5	AAU99339	Aau99339 Clostridi
33	772.5	30.3	660	4	AAE07898	Aae07898 Modified
34	763	29.9	451	2	AAW68398	Aaw68398 Clostridi
35	762	29.9	415	4	AAB04083	Aab04083 Botulism
36	743	29.1	449	3	AAY77137	Aay77137 Synthetic
37	743	29.1	449	4	AAB04094	Aab04094 Botulism
38	738	28.9	419	4	AAB04095	Aab04095 Botulism
39	729.5	28.6	685	4	AAE07893	Aae07893 Modified
40	727	28.5	862	4	AAE07890	Aae07890 Modified
41	727	28.5	887	4	AAE07892	Aae07892 Modified
42	727	28.5	979	6	AAE35713	Aae35713 BoNT/F-Hc
43	727	28.5	1192	6	AAE35711	Aae35711 BoNT/F-Hc
44	727	28.5	1192	6	AAE35710	Aae35710 BoNT/F-Hc
45	726.5	28.5	1092	4	AAE07900	Aae07900 C. botuli

ALIGNMENTS

RESULT 1
AAW68400
ID AAW68400 standard; protein; 473 AA.
XX
AC AAW68400;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type G toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
KW botulism; BotG.
XX
OS Clostridium botulinum; serotype G strain 113/30.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30596.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (113/30 strain) type G neurotoxin, encoded by a DNA
CC sequence (see AAV30596) in plasmid pETHisb. This vector can be used to
CC express BotG soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on an affinity column. The
CC invention relates to recombinant proteins derived from C. botulinum
CC toxins, especially type B and type E toxins. Methods are provided which

```
CC   allow for the isolation of soluble recombinant proteins free of
CC   significant endotoxin contamination. Preferred hosts for production of
CC   recombinant proteins are E. coli, insect cells and yeast cells. The
CC   recombinant toxins are used as immunogens for the production of vaccines
CC   and antitoxins that are useful in the treatment of humans and animals at
CC   risk of intoxication with clostridial toxin
XX
SQ   Sequence 473 AA;

  Query Match            100.0%;  Score 2550;  DB 2;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 1e-182;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 2
AAY77139
ID   AAY77139 standard; protein; 449 AA.
XX
AC   AAY77139;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype G (BoNTG) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype G; C-terminal fragment;
KW   Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW   diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US015570.
XX
PR   10-JUL-1998;   98US-0092416P.
PR   12-MAY-1999;   99US-0133870P.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87217.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum.
XX
PS   Claim 28; Page 47-48; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid fragment
CC   comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC   G. In preferred embodiments of the invention, the vector is a Venezuelan
CC   equine encephalitis virus (VEE) replicon vector. Use of this vector
CC   results in the production of large amounts of a protein encoded by a
CC   sequence cloned into the replicon. The constructs are used to produce
CC   vaccines against botulism. The proteins can also be used as diagnostic
CC   tools for the diagnosis of botulism. The transformed host cells can be
CC   used to analyse the effectiveness of drugs and agents which inhibit toxin
CC   effects. The vaccine currently used against botulism is dangerous and
CC   expensive to produce, and contains formalin, which is very painful for
CC   the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC   serotypes are represented in the formulation. The novel vaccine of
CC   overcomes these problems, as it is easily purified, and available in
CC   large quantities. It is also expressed in the lymph nodes for a better
CC   immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC   fragments used in the present invention. The DNA encoding these sequences
CC   had been optimised for codon usage for expression in yeast
XX
SQ   Sequence 449 AA;

  Query Match            93.8%;  Score 2391;  DB 3;  Length 449;
  Best Local Similarity  99.8%;  Pred. No. 8.1e-171;
  Matches  448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         25 MADTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 84
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         85 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        205 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        265 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 300

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 360

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 420

Qy        445 YLRRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db        421 YLRRISENINKLRLGCNWQFIPVDEGWTE 449
```

```
RESULT 3
AAB04167
ID   AAB04167 standard; protein; 449 AA.
XX
AC   AAB04167;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype G).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
PR   12-MAY-1999;   99US-0133868P.
PR   12-MAY-1999;   99US-0133869P.
PR   12-MAY-1999;   99US-0133873P.
PR   29-JUL-1999;   99US-0146192P.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54491.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT   heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT   against botulism.
XX
PS   Claim 3; Fig 10b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC   and then posttranslationally nicked, forming a dichain consisting of a
CC   100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC   disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC   -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC   can be used in recombinant expression vectors and expressed in
CC   transformed cells to produce peptide antigens useful for eliciting an
CC   immune response to give protective immunity against botulinum neurotoxin,
CC   which causes botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use of
CC   recombinant nucleic acids are advantageous since it eliminates the need
CC   to culture large quantities of hazardous toxin-producing bacterium.
CC   Production yield from the genetically engineered product is also high and
CC   cost of production is lower. The nucleic acids can be derived from
CC   Clostridium botulinum serotypes A-G
XX
SQ   Sequence 449 AA;

  Query Match             93.8%;  Score 2391;  DB 4;  Length 449;
  Best Local Similarity   99.8%;  Pred. No. 8.1e-171;
  Matches  448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         25 MADTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 84
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         85 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        205 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        265 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 300

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 360

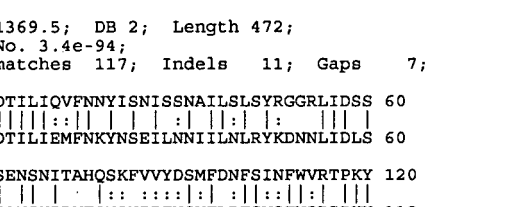
Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 420

Qy        445 YLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||
Db        421 YLRRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 4
AAW68394
ID   AAW68394 standard; protein; 472 AA.
XX
AC   AAW68394;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum toxin B C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30581.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 303-305; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (Danish strain) type B toxin, encoded by a DNA
CC   sequence (see AAV30581) in plasmid pETHisb. This vector was used to
CC   express soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins. Methods
```

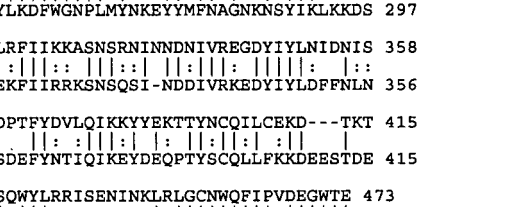
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 53.7%; Score 1369.5; DB 2; Length 472;
Best Local Similarity 53.3%; Pred. No. 3.4e-94;
Matches 255; Conservative 95; Mismatches 117; Indels 11; Gaps 7;



RESULT 5
AAW68393
ID AAW68393 standard; protein; 472 AA.
XX
AC AAW68393;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin B C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB.
XX
OS Clostridium botulinum; serotype B strain Eklund 17B.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30580.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment of
CC Clostridium botulinum (Eklund 17B strain) type B toxin, encoded by a DNA
CC sequence (see AAV30580) in plasmid pHisBotb. This vector was used to
CC express soluble C fragment in Escherichia coli host cells, and the
CC recombinant C fragment was purified on an affinity column. The invention
CC relates to recombinant proteins derived from C. botulinum toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred hosts for
CC production of recombinant proteins are E. coli, insect cells and yeast
CC cells. The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of humans and
CC animals at risk of intoxication with clostridial toxin
XX
SQ Sequence 472 AA;

Query Match 53.0%; Score 1350.5; DB 2; Length 472;
Best Local Similarity 52.9%; Pred. No. 9.1e-93;
Matches 253; Conservative 94; Mismatches 120; Indels 11; Gaps 7;



RESULT 6
AAW68392
ID AAW68392 standard; protein; 1291 AA.
XX
AC AAW68392;
XX
DT 17-OCT-2003 (revised)

```
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type B toxin.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW   botulism; BotB.
XX
OS   Clostridium botulinum; serotype B Danish strain.
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA, Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30579.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 35; Page 291-296; 428pp; English.
XX
CC   This is the amino acid sequence of the type B toxin of Clostridium
CC   botulinum Danish strain. The C fragment (see AAW68393-94) of the serotype
CC   B toxin has been expressed as a histidine-tagged protein in Escherichia
CC   coli. The invention relates to C. botulinum recombinant toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred hosts for
CC   production of the recombinant proteins are E. coli, insect cells and
CC   yeast cells. The recombinant proteins are used as immunogens for the
CC   production of vaccines and antitoxins that are useful in the treatment of
CC   humans and animals at risk of intoxication with clostridial toxin.
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1291 AA;

  Query Match             47.5%;  Score 1211.5;  DB 2;  Length 1291;
  Best Local Similarity   50.7%;  Pred. No. 9.3e-82;
  Matches  229;  Conservative   96;  Mismatches  116;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              |||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||    |||| :
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 903

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: :::::|:| :||::||:| ||| |: || |: ||||||:|:||:||||
Db        904 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        964 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 1022

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              : :  | ::  : :: :| |||     | |:|:|:| |:||  ||: : :   | ||| :
Db       1023 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 1082

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||||  |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db       1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 1142

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              .:|||:: |||::| ||:|||: |||||:  |:: | :|||    .|: : :||||||:|
Db       1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 1200

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
                 ||: :|||:|:|: ||:||:| :||   |   || || :| :  | |:: | :||||
Db       1201 SDEFYNTIQIKEYHEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 1259

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :      |:|||||||| ||||||
Db       1260 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 7
AAY93308
ID   AAY93308 standard; protein; 1070 AA.
XX
AC   AAY93308;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype B.
XX
OS   Synthetic.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 45-47; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that can
CC   be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype B. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a domain
CC   that translocates the SOD of the composition into the neuronal cell.
CC   After translocation, the linker is cleaved to release the SOD. The
CC   composition is useful for treating neuronal diseases caused or augmented
CC   by oxidative stress, such as ischemic stroke, trauma, Parkinson's
CC   disease, Huntington's disease and motor neurone diseases
XX
SQ   Sequence 1070 AA;

  Query Match             47.3%;  Score 1206.5;  DB 3;  Length 1070;
  Best Local Similarity   50.7%;  Pred. No. 1.7e-81;
  Matches  229;  Conservative   95;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              |||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||    |||| :
Db        625 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 682

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: :::::|:| :||::||:| ||| |: || |: ||||||:|:||:||||
```

```
Db        683 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 742

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        743 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 801

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              : :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :
Db        802 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 861

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||||  |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db        862 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 921

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||:|
Db        922 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 979

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
                ||: :|||:| |: ||:||:| :||    |   || || :| :  | |:: | :||||
Db        980 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 1038

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :      |:|||||||| ||||||
Db       1039 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1070


RESULT 8
AAY93311
ID   AAY93311 standard; protein; 1095 AA.
XX
AC   AAY93311;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype B.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Geobacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB003699.
XX
PR   05-NOV-1998;   98GB-00024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a cleavable
PT   linker to a neuronal cell targeting component useful for delivering
PT   superoxide dismutase to neuronal cells to treat ischemia.
XX
PS   Disclosure; Page 54-57; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide dismutase
CC   (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker that can be
CC   cleaved by thrombin, and a heavy chain derived from botulinum neurotoxin
CC   serotype B. The specification describes a composition for delivery of SOD
CC   to neuronal cells. The composition comprises SOD linked, by a cleavable
CC   linker, to a neuronal cell targeting component (NCTC). This component has
CC   a domain that binds to a neuronal cell and a domain that translocates the
CC   SOD of the composition into the neuronal cell. After translocation, the
CC   linker is cleaved to release the SOD. The composition is useful for
CC   treating neuronal diseases caused or augmented by oxidative stress, such
CC   as ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases
XX
SQ   Sequence 1095 AA;

  Query Match             47.3%;  Score 1206.5;  DB 3;  Length 1095;
  Best Local Similarity   50.7%;  Pred. No. 1.8e-81;
  Matches  229;  Conservative   95;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              |||||::|| | | | :| ||:| |:   ||| ||||| : |   | || . |||| :
Db        650 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 707

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:|||
Db        708 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 767

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        768 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 826

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              : :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :
Db        827 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 886

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||||  |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db        887 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 946

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||:|
Db        947 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 1004

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
                ||: :|||:| |: ||:||:| :||    |   || || :| :  | |:: | :||||
Db       1005 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 1063

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :      |:|||||||| ||||||
Db       1064 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1095


RESULT 9
AAB04085
ID   AAB04085 standard; protein; 439 AA.
XX
AC   AAB04085;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin C fragment sequence (serotype B).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US012890.
XX
PR   12-MAY-1999;   99US-0133865P.
PR   12-MAY-1999;   99US-0133866P.
PR   12-MAY-1999;   99US-0133867P.
```

PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Example 8; Page 38; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 439 AA;

Query Match 46.1%; Score 1175.5; DB 4; Length 439;
Best Local Similarity 50.1%; Pred. No. 1.1e-79;
Matches 223; Conservative 92; Mismatches 119; Indels 11; Gaps 7;

Qy 34 FNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNIT 93
Db 1 FNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSANSKIR 58

Qy 94 AHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNR 153
Db 59 VTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNR 118

Qy 154 IIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKIL 213
Db 119 ITWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIK 177

Qy 214 NLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWG 273
Db 178 DIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWG 237

Qy 274 NPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLRFIIKK 331
Db 238 NPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRR 297

Qy 332 ASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDV 391
Db 298 KSNSQSI-NDDIVRKEDYIYLDFFNLNQEQ-RVYTYKKFKKEEEKLFLAPISDSDEFYNT 355

Qy 392 LQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRR 448
Db 356 IQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCISKWYLKE 414

Qy 449 ISENINKLRLGCNWQFIPVDEGWTE 473
Db 415 VKRKPYNLKLGCNWQFIPKDEGWTE 439

RESULT 10

AAY77135
ID AAY77135 standard; protein; 440 AA.
XX
AC AAY77135;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype B (BoNTB) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype B; C-terminal fragment;
KW Venezuelan equine encephalitis virus replicon; VEE; botulism; vaccine;
KW diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US015570.
XX
PR 10-JUL-1998; 98US-0092416P.
PR 12-MAY-1999; 99US-0133870P.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87213.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum.
XX
PS Claim 23; Page; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid fragment
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-
CC G. In preferred embodiments of the invention, the vector is a Venezuelan
CC equine encephalitis virus (VEE) replicon vector. Use of this vector
CC results in the production of large amounts of a protein encoded by a
CC sequence cloned into the replicon. The constructs are used to produce
CC vaccines against botulism. The proteins can also be used as diagnostic
CC tools for the diagnosis of botulism. The transformed host cells can be
CC used to analyse the effectiveness of drugs and agents which inhibit toxin
CC effects. The vaccine currently used against botulism is dangerous and
CC expensive to produce, and contains formalin, which is very painful for
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7
CC serotypes are represented in the formulation. The novel vaccine of
CC overcomes these problems, as it is easily purified, and available in
CC large quantities. It is also expressed in the lymph nodes for a better
CC immune response. Sequences AAY77134-Y77139 represent synthetic BoNT Hc
CC fragments used in the present invention. The DNA encoding these sequences
CC had been optimised for codon usage for expression in yeast. Note: This
CC sequence is not given in the specification, but is decoded from the BoNTB
CC Hc DNA sequence given on pages 39-40
XX
SQ Sequence 440 AA;

Query Match 45.8%; Score 1168.5; DB 3; Length 440;
Best Local Similarity 50.0%; Pred. No. 3.7e-79;
Matches 222; Conservative 93; Mismatches 118; Indels 11; Gaps 7;

Qy 35 NNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITA 94
Db 3 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSANSKIRV 60

Qy 95 HQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRI 154

```
                |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:||||:||:||||
Db         61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILN 214
              ||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |: :  | :
Db        121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKD 179

Qy        215 LDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGN 274
              :  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :  |||||||
Db        180 IREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGN 239

Qy        275 PLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLRFIIKKA 332
              || |: :||:|| | :| |||    : :||   |: :|  :  |||::||:| :|||::
Db        240 PLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRK 299

Qy        333 SNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVL 392
              |||::| ||:|||: |||||:  |:: | :|||      |: : :||||||:|    |: :
Db        300 SNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISDSDELYNTI 357

Qy        393 QIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRI 449
              |||:| |: ||:||:| :||   |   || || :| :  | |:: | :|||||:|||: :
Db        358 QIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCISKWYLKEV 416

Qy        450 SENINKLRLGCNWQFIPVDEGWTE 473
                    |:|||||||| ||||||
Db        417 KRKPYNLKLGCNWQFIPKDEGWTE 440
```

RESULT 11
AAB04091
ID AAB04091 standard; protein; 440 AA.
XX
AC AAB04091;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54485.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Claim 3; Fig 4b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 440 AA;

  Query Match   45.8%; Score 1168.5; DB 4; Length 440;
  Best Local Similarity   50.0%; Pred. No. 3.7e-79;
  Matches 222; Conservative   93; Mismatches 118; Indels   11; Gaps   7;

```
Qy         35 NNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITA 94
              | | | | :| ||:| |:   ||| ||||| : |   |  ||    |||| :| || |
Db          3 NKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTSSANSKIRV 60

Qy         95 HQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRI 154
               |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:||||:||:||||
Db         61 TQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRI 120

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILN 214
              ||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |: :  | :
Db        121 IWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKLESNTDIKD 179

Qy        215 LDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGN 274
              :  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :  |||||||
Db        180 IREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGN 239

Qy        275 PLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLRFIIKKA 332
              || |: :||:|| | :| |||    : :||   |: :|  :  |||::||:| :|||::
Db        240 PLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIGEKFIIRRK 299

Qy        333 SNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVL 392
              |||::| ||:|||: |||||:  |:: | :|||      |: : :||||||:|    |: :
Db        300 SNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISDSDELYNTI 357

Qy        393 QIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRI 449
              |||:| |: ||:||:| :||   |   || || :| :  | |:: | :|||||:|||: :
Db        358 QIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCISKWYLKEV 416

Qy        450 SENINKLRLGCNWQFIPVDEGWTE 473
                    |:|||||||| ||||||
Db        417 KRKPYNLKLGCNWQFIPKDEGWTE 440
```

RESULT 12
AAB04082
ID AAB04082 standard; protein; 848 AA.
XX
AC AAB04082;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain (serotype B).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.

XX
PF 12-MAY-2000; 2000WO-US012890.
XX
PR 12-MAY-1999; 99US-0133865P.
PR 12-MAY-1999; 99US-0133866P.
PR 12-MAY-1999; 99US-0133867P.
PR 12-MAY-1999; 99US-0133868P.
PR 12-MAY-1999; 99US-0133869P.
PR 12-MAY-1999; 99US-0133873P.
PR 29-JUL-1999; 99US-0146192P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of the
PT heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine
PT against botulism.
XX
PS Disclosure; Page 13; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide chain
CC and then posttranslationally nicked, forming a dichain consisting of a
CC 100 kDa heavy chain and a 50 kDa light chain which remain linked by a
CC disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino
CC -terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT)
CC can be used in recombinant expression vectors and expressed in
CC transformed cells to produce peptide antigens useful for eliciting an
CC immune response to give protective immunity against botulinum neurotoxin,
CC which causes botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use of
CC recombinant nucleic acids are advantageous since it eliminates the need
CC to culture large quantities of hazardous toxin-producing bacterium.
CC Production yield from the genetically engineered product is also high and
CC cost of production is lower. The nucleic acids can be derived from
CC Clostridium botulinum serotypes A-G
XX
SQ Sequence 848 AA;

Query Match 44.8%; Score 1143.5; DB 4; Length 848;
Best Local Similarity 49.9%; Pred. No. 6.6e-77;
Matches 226; Conservative 93; Mismatches 119; Indels 15; Gaps 11;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              |||||::|| | | :| ||:| |:   ||| ||||| : |   | ||    |||| :
Db        405 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 462

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: :::::|:| :||::||:| ||| |: || |: ||||||:|:||:|||| 
Db        463 SANSKIRTTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 522

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
             :||:| |||||||:| |:||:||||:|:::||:|||:|| :|||: | || ||||| |
Db        523 ISIRG-RIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 580

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
             : :  | ::  : :: :| |||     | |:|:|:| |:||  ||: : :   | ||| :
Db        581 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 640

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIY-IKYFSKASMGETAPRTNFNNAA--INYQNLYL 323
             |||||||||| |: :||:|| | :| | ||    : :||   |: :|  :  |||::||:
Db        641 YLKDFWGNPLMYNKEYYMFNAGNKNSYTIKLKKDSPVGEILTRSKYNQNSKYINYRDLYI 700

Qy        324 GLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIN 383
             | :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||:
Db        701 GEKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKK-EEELFLAPIS 757

Qy        384 DDPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFC 440
             |   ||: :|||:| |: ||:||:| :||    |   || || :| :  | |::   : ||
Db        758 DSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEVKDVFC 816
```

Qy 441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
|| ||| : |:|||||||| ||||||
Db 817 ISDWYL-EVKRKPYNLKLGCNWQFIPKDEGWTE 848

RESULT 13
AAR95009
ID AAR95009 standard; protein; 462 AA.
XX
AC AAR95009;
XX
DT 07-JUL-1996 (first entry)
XX
DE Type A neurotoxin C fragment-polyhistidine tag fusion pHisBot.
XX
KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KW Clostridium botulinum; polyhistidine; vector; pETHisa; pHisBot.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /label= Polyhistidine_tag
FT Protein 22. .462
FT /label= C-fragment
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29246.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficile type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 7; Page 340-342; 434pp; English.
XX
CC pHisBot fusion protein (AAR95009), the product of a nucleotide sequence
CC (AAT29246) in vector pETHisa, comprises a polyhistidine affinity tag and
CC fragment C (see also AAR95008) of the Clostridium botulinum type A
CC neurotoxin. The pHisBot protein was expressed in Escherichia coli as a
CC soluble protein and was purified by metal chelate affinity chromatography
CC to obtain a product free of endotoxin contamination that may be useful as
CC an immunogen in vaccine compsns
XX
SQ Sequence 462 AA;

Query Match 37.8%; Score 965; DB 2; Length 462;
Best Local Similarity 42.9%; Pred. No. 7.1e-64;
Matches 207; Conservative 77; Mismatches 166; Indels 32; Gaps 12;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              |||||||||||||||||||||||||||||||   |:  |  || || : :||:| |    ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
               | : :|:|| | |: |   | :| | |:| |      :||:||::||| :||:| |||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118
```

```
Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
Db        119 FNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-- 298
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        299 -KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
Db        296 IRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDNIVRNNDRVYINV- 350

Qy        356 NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKTTYNCQI-LCEKDT 413
Db        351 VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNG 410

Qy        414 KTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRLGCNWQFIPVDEGW 471
Db        411 NDIGFIGFHQF-----------NNIAKLVASNWYNRQIER--SSRTLGCSWEFIPVDDGW 457

Qy        472 TE 473
Db        458 GE 459
```

RESULT 14
AAW68390
ID AAW68390 standard; protein; 462 AA.
XX
AC AAW68390;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C (His-tagged).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism.
XX
OS Clostridium botulinum; serotype A.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .21
FT /note= "N-terminal histidine tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30572, AAV30575.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 24; Page 265-267; 428pp; English.
XX
CC This is the amino acid sequence of a histidine-tagged fragment C
CC polypeptide of Clostridium botulinum serotype A toxin encoded by a DNA
CC sequence (see AAV30572) in plasmid pHisBot, and by a DNA sequence (see
CC AAV30575) in pHisBotA. These vectors were used to express native (i.e.
CC non-fusion) soluble C fragment in Escherichia coli host cells, with the
CC recombinant C fragment being purified on a poly-histidine binding
CC affinity resin. The invention relates to recombinant proteins derived
CC from C. botulinum toxins. Methods are provided which allow the isolation
CC of soluble recombinant proteins that are free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant proteins are
CC E. coli, insect cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are useful
CC in the treatment of humans and animals at risk of intoxication with
CC clostridial toxin
XX
SQ Sequence 462 AA;

Query Match 37.8%; Score 965; DB 2; Length 462;
Best Local Similarity 42.9%; Pred. No. 7.1e-64;
Matches 207; Conservative 77; Mismatches 166; Indels 32; Gaps 12;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
Db        119 FNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-- 298
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        299 -KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
Db        296 IRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDNIVRNNDRVYINV- 350

Qy        356 NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKTTYNCQI-LCEKDT 413
Db        351 VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNG 410

Qy        414 KTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRLGCNWQFIPVDEGW 471
Db        411 NDIGFIGFHQF-----------NNIAKLVASNWYNRQIER--SSRTLGCSWEFIPVDDGW 457

Qy        472 TE 473
Db        458 GE 459
```

RESULT 15
AAW68396
ID AAW68396 standard; protein; 452 AA.
XX
AC AAW68396;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin E C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE.
XX
OS Clostridium botulinum; serotype E strain NCTC 11219.
OS Synthetic.
XX
FH Key Location/Qualifiers

```
FT   Peptide          1. .21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US015394.
XX
PR   28-AUG-1996;   96US-00704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Williams JA,  Thalley BS;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30585.
XX
PT   Host cell containing recombinant expression vector encoding Clostridium
PT   botulinum type B or E toxin - useful to treat humans and other animals at
PT   risk of intoxication with clostridial toxin.
XX
PS   Example 41; Page 327-329; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment of
CC   Clostridium botulinum (NCTC 11219) type E neurotoxin, encoded by a DNA
CC   sequence (see AAV30585) in plasmid pETHisb. This vector is used to
CC   express BotE soluble C fragment in Escherichia coli host cells, and the
CC   recombinant C fragment was purified on an affinity column. The invention
CC   relates to recombinant proteins derived from C. botulinum toxins,
CC   especially type B and type E toxins. Methods are provided which allow for
CC   the isolation of soluble recombinant proteins free of significant
CC   endotoxin contamination. Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells. The recombinant
CC   toxins are used as immunogens for the production of vaccines and
CC   antitoxins that are useful in the treatment of humans and animals at risk
CC   of intoxication with clostridial toxin
XX
SQ   Sequence 452 AA;

  Query Match           35.3%;  Score 900;  DB 2;  Length 452;
  Best Local Similarity 39.2%;  Pred. No. 5.2e-59;
  Matches  194;  Conservative   81;  Mismatches  154;  Indels   66;  Gaps   15;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMA------DTILIQVFNNYISNISSNAILSLSYRGG 54
              ||||||||||||||||||||||||||      | |||  || : . | |:::|:: |:
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMALSSYTDDKILISYFNKFFKRIKSSSVLNMRYKND 60

Qy         55 RLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFW 114
             : :|:||| : :|:  ||        || ::| : | :   |:::::||: : ||||:||
Db         61 KYVDTSGYDSNININGDVYKYPTNKNQFGIDNDKLSEVNISQNDYIIYDNKYKNFSISFW 120

Qy        115 VRTPKYNNNDIQTYLQNEYTIISCIK-NDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYS 173
              || | |:|  :   : ||||||:|:: |:|||||:  | ||||| |    :: : | |
Db        121 VRIPNYDNKIVN--VNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYG 178

Qy        174 IKDNISDYINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTD 233
                : ||||||||  :||||||||:: :||||:|  : ||||  |: |::| ||::||:
Db        179 NANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCS- 237

Qy        234 TTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIY 293
              |::: |: |||| :||: ||: :||  : :|| ||||||| | || :||| |    | :
Db        238 YTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNF 297

Qy        294 IKYFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLN 353
              |          |    |  :  :    || |::  |:: :||    |||:||: | :|.| .
Db        298 I----DRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSS--TNDNLVRKNDQVYIN 351

Qy        354 IDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKT--------TYNC 405
                        | ||   | ||       | : |     |  |||         :|
Db        352 ------------FVASK---THLF-------PLYADTATTNK--EKTIKISSSGNRFNQ 386

Qy        406 QIL-------CEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRL 458
```

```
                ::       |  : |      ||      |: ||       | ||   : :: |
Db        387 VVVMNSVGNNCTMNFKNNNGNNIGLL----GFKADT----VVASTWYYTHMRDHTNS--N 436

Qy        459 GCNWQFIPVDEGWTE 473
              || | ||  : || |
Db        437 GCFWNFISEEHGWQE 451
```

Search completed: September 10, 2004, 03:33:48
Job time : 33.2007 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:57 ; Search time 10.0395 Seconds
(without alignments)
2432.294 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 2550
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	965	37.8	462	1	US-08-480-604A-26	Sequence 26, Appl
2	965	37.8	462	2	US-08-405-496A-26	Sequence 26, Appl
3	965	37.8	462	3	US-08-915-136-26	Sequence 26, Appl
4	965	37.8	462	4	US-09-084-517-26	Sequence 26, Appl
5	894	35.1	1169	4	US-09-255-829-20	Sequence 20, Appl
6	815	32.0	438	1	US-08-480-604A-23	Sequence 23, Appl
7	815	32.0	438	2	US-08-405-496A-23	Sequence 23, Appl
8	815	32.0	438	3	US-08-915-136-23	Sequence 23, Appl
9	815	32.0	438	4	US-09-084-517-23	Sequence 23, Appl
10	815	32.0	1296	1	US-08-480-604A-28	Sequence 28, Appl
11	815	32.0	1296	2	US-08-405-496A-28	Sequence 28, Appl
12	815	32.0	1296	3	US-08-915-136-28	Sequence 28, Appl
13	815	32.0	1296	4	US-09-084-517-28	Sequence 28, Appl
14	643.5	25.2	452	1	US-07-618-312A-2	Sequence 2, Appli
15	643.5	25.2	452	1	US-08-280-228-2	Sequence 2, Appli
16	642.5	25.2	452	1	US-07-618-312A-4	Sequence 4, Appli
17	642.5	25.2	452	1	US-08-280-228-4	Sequence 4, Appli
18	642.5	25.2	618	1	US-08-668-381A-5	Sequence 5, Appli
19	642.5	25.2	853	4	US-08-913-880C-17	Sequence 17, Appl
20	642.5	25.2	858	4	US-08-913-880C-16	Sequence 16, Appl
21	642.5	25.2	860	4	US-08-913-880C-15	Sequence 15, Appl
22	642.5	25.2	862	4	US-08-913-880C-14	Sequence 14, Appl
23	642.5	25.2	865	4	US-08-913-880C-13	Sequence 13, Appl
24	642.5	25.2	866	4	US-08-913-880C-12	Sequence 12, Appl
25	642.5	25.2	874	4	US-08-913-880C-11	Sequence 11, Appl
26	642.5	25.2	875	4	US-08-913-880C-10	Sequence 10, Appl
27	642.5	25.2	1315	4	US-08-913-880C-1	Sequence 1, Appli
28	618	24.2	452	1	US-08-110-786A-8	Sequence 8, Appli
29	164.5	6.5	141	4	US-09-465-276-1	Sequence 1, Appli
30	158.5	6.2	140	3	US-08-446-114A-22	Sequence 22, Appl
31	158	6.2	2710	1	US-08-480-604A-6	Sequence 6, Appli
32	158	6.2	2710	2	US-08-405-496A-6	Sequence 6, Appli
33	158	6.2	2710	3	US-08-915-136-6	Sequence 6, Appli
34	158	6.2	2710	4	US-08-957-310-6	Sequence 6, Appli
35	158	6.2	2710	4	US-10-011-366-6	Sequence 6, Appli
36	158	6.2	2710	4	US-09-084-517-6	Sequence 6, Appli
37	150	5.9	23	1	US-08-480-604A-24	Sequence 24, Appl
38	150	5.9	23	2	US-08-405-496A-24	Sequence 24, Appl
39	150	5.9	23	3	US-08-915-136-24	Sequence 24, Appl
40	150	5.9	23	4	US-09-084-517-24	Sequence 24, Appl
41	147	5.8	323	1	US-08-591-989-5	Sequence 5, Appli
42	146	5.7	22	4	US-09-660-742-3	Sequence 3, Appli
43	146	5.7	53	2	US-08-651-818A-19	Sequence 19, Appl
44	146	5.7	53	3	US-09-184-826-19	Sequence 19, Appl
45	146	5.7	54	2	US-08-651-818A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-26
; Sequence 26, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-26

  Query Match             37.8%;  Score 965;  DB 1;  Length 462;
  Best Local Similarity   42.9%;  Pred. No. 1.1e-68;
  Matches  207;  Conservative   77;  Mismatches  166;  Indels   32;  Gaps   12;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||    |:  |  || || : :||:| |    ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
                | : :|:|| | |: |    | :| | |:| |        :||:||::||| :||:| |||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
               |:     | ||||||:|::|:||||||:    |||||  |     : : |:||   ||||
Db        119 FNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
               |||:|  :||||:|| |: ||||| |   : | ||  |::||:| |||  | || :::||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-- 298
               | ||:| :|||  |:  ||  ||::  ||||||: |:||  ||: |    | |:   :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        299 -KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
              :  |    || :    |  :|  :|| | :|||| ::    | |||||  | :|:|:
Db        296 IRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDNIVRNNDRVYINV- 350

Qy        356 NISDESYRVYYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKTTYNCQI-LCEKDT 413
              : :: ||:    :  ::  |    | |      |: :| |  :  |  |:: | : :
Db        351 VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNG 410

Qy        414 KTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRLGCNWQFIPVDEGW 471
                | |  :|          :|      | || |:|   :  |||:|:|||||:||
Db        411 NDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTLGCSWEFIPVDDGW 457

Qy        472 TE 473
               |
Db        458 GE 459


RESULT 2
US-08-405-496A-26
; Sequence 26, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-26

  Query Match             37.8%;  Score 965;  DB 2;  Length 462;
  Best Local Similarity   42.9%;  Pred. No. 1.1e-68;
  Matches  207;  Conservative   77;  Mismatches  166;  Indels   32;  Gaps   12;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||    |:  |  || || : :||:| |    ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
                | : :|:|| | |: |    | :| | |:| |        :||:||::||| :||:| |||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
               |:     | ||||||:|::|:||||||:    |||||  |     : : |:||   ||||
Db        119 FNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
               |||:|  :||||:|| |: ||||| |   : | ||  |::||:| |||  | || :::||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-- 298
               | ||:| :|||  |:  ||  ||::  ||||||: |:||  ||: |    | |:   :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        299 -KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
              :  |    || :    |  :|  :|| | :|||| ::    | |||||  | :|:|:
Db        296 IRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDNIVRNNDRVYINV- 350

Qy        356 NISDESYRVYYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKTTYNCQI-LCEKDT 413
              : :: ||:    :  ::  |    | |      |: :| |  :  |  |:: | : :
Db        351 VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNG 410

Qy        414 KTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRLGCNWQFIPVDEGW 471
                | |  :|          :|      | || |:|   :  |||:|:|||||:||
Db        411 NDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTLGCSWEFIPVDDGW 457

Qy        472 TE 473
               |
Db        458 GE 459
```

```
RESULT 3
US-08-915-136-26
; Sequence 26, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-26

  Query Match             37.8%;  Score 965;  DB 3;  Length 462;
  Best Local Similarity   42.9%;  Pred. No. 1.1e-68;
  Matches  207;  Conservative   77;  Mismatches  166;  Indels   32;  Gaps   12;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||    |:  |  || || : :||:| |    ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              | : :|:|| | |: |   | :| | |:| |       :||:||::||| :||:| |||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              |:     | ||||||:|::|:||||||:    ||||| |     : : |:||   ||||
Db        119 FNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              |||:|  :||||:|| |: ||||| |   : | ||  |::||:| |||  | || :::||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-- 298
              | ||:| :|||  |:  ||  ||::  ||||||: |:||  ||: |    | |:   :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        299 -KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
               :  |    || :    | :|  :|| | :||||| ::    | |||||  | :|:|:
Db        296 IRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDNIVRNNDRVYINV- 350

Qy        356 NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKTTYNCQI-LCEKDT 413
               : :: ||:    :   ::  |    | |      |: :| |  :  |  |:: | : :
Db        351 VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNG 410

Qy        414 KTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRLGCNWQFIPVDEGW 471
              |  | :|           :|      | || |:|    :   |||:|:|||||:||
Db        411 NDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTLGCSWEFIPVDDGW 457

Qy        472 TE 473
              |
Db        458 GE 459


RESULT 4
US-09-084-517-26
; Sequence 26, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
```

; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-26

Query Match 37.8%; Score 965; DB 4; Length 462;
Best Local Similarity 42.9%; Pred. No. 1.1e-68;
Matches 207; Conservative 77; Mismatches 166; Indels 32; Gaps 12;

```
Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||   |:  |  || || : :||:| |    ||| |
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMAR--LLSTFTEYIKNIINTSILNLRYESNHLIDLS 58

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
               | : :|:|| | |: |   | :| | |:| |       :||:||::||| :||:| |||
Db         59 RYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKY 118

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              |:     | ||||||:|::|:||||||:    ||||| |     : : |:||   ||||
Db        119 FNS---ISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISD 175

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              |||:|  :||||:|| |: ||||| |   : | ||  |::||:| |||  | || :::||
Db        176 YINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWI 235

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-- 298
              | ||:| :|||  |:  ||  ||::  ||||||: |:|| ||: |   | |:   :
Db        236 KYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVG 295

Qy        299 -KASMGETAPRTNF--NNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
               :  |    || :   |  :|  :|| | :||||| ::    | |||||  | :|:|:
Db        296 IRGYMYLKGPRGSVMTTNIYLN-SSLYRGTKFIIKKYASG---NKDNIVRNNDRVYINV- 350

Qy        356 NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK-KYYEKTTYNCQI-LCEKDT 413
               : :: ||:    :   ::  |    | |       |: :| |  :  |  |:: | : :
Db        351 VVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNG 410

Qy        414 KTFGLFGIGKFVKDYGYVWDTYDNY--FCISQWYLRRISENINKLRLGCNWQFIPVDEGW 471
              |  |  :|          :|      | || |:|    :   |||:|:||||:||
Db        411 NDIGFIGFHQF----------NNIAKLVASNWYNRQIER--SSRTLGCSWEFIPVDDGW 457

Qy        472 TE 473
              ||
Db        458 GE 459
```

RESULT 5
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-20

Query Match 35.1%; Score 894; DB 4; Length 1169;
Best Local Similarity 51.5%; Pred. No. 1.7e-62;
Matches 169; Conservative 71; Mismatches 82; Indels 6; Gaps 4;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              |||||::|| | | | :| ||:| |:   ||| ||||| : |   |  ||   |||| :
Db        846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 903

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               | || |   |:: ::::|:| :||::||:| ||| |: || |: |||||:|:||:||||
Db        904 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        964 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 1022

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              : :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :   | ||| :
Db       1023 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 1082

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||| |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db       1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 1142

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYL 352
               :|||:: |||::| ||:|||: |||||
Db       1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYL 1169
```

RESULT 6
US-08-480-604A-23
; Sequence 23, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO

```
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-23

  Query Match           32.0%;  Score 815;  DB 1;  Length 438;
  Best Local Similarity  40.0%;  Pred. No. 8.5e-57;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |    | :| | |:
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 120

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
              ||||| |     : : |:||   |||||||:|  :||||:|| |: ||||| |
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |    | |:   :   : |    || :    |  :|  :||'|
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 299

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :||||| ::    | |||||  | :|:|:  : :: ||:    :  ::  |    | |
Db        300 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 355

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
                       |: :| |  :  |  |:: | : :    |  |  :|           :|
Db        356 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 404

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
                | || |:|    :   |||:|:|||||:|| |
Db        405 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 435


RESULT 7
US-08-405-496A-23
; Sequence 23, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
;   APPLICANT:  WILLIAMS, JAMES A.
;   TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;   TITLE OF INVENTION:  NEUROTOXIN
;   NUMBER OF SEQUENCES:  30
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  MEDLEN & CARROLL, LLP
;     STREET:  220 MONTGOMERY STREET, SUITE 2200
;     CITY:  SAN FRANCISCO
;     STATE:  CALIFORNIA
;     COUNTRY:  USA
;     ZIP:  94104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/405,496A
;     FILING DATE:  16-MAR-1995
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/329,154
;     FILING DATE:  25-OCT-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/161,907
;     FILING DATE:  02-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/985,321
;     FILING DATE:  04-DEC-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/429,791
;     FILING DATE:  31-OCT-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  INGOLIA, DIANE E.
;     REGISTRATION NUMBER:  40,027
;     REFERENCE/DOCKET NUMBER:  OPHD-01308
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 705-8410
;     TELEFAX:  (415) 397-8338
; INFORMATION FOR SEQ ID NO:  23:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  438 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-405-496A-23

  Query Match           32.0%;  Score 815;  DB 2;  Length 438;
  Best Local Similarity  40.0%;  Pred. No. 8.5e-57;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |    | :| | |:
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 120
```

```
Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 299

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
Db        300 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 355

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
Db        356 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 404

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db        405 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 435


RESULT 8
US-08-915-136-23
; Sequence 23, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-915-136-23

  Query Match           32.0%;  Score 815;  DB 3;  Length 438;
  Best Local Similarity  40.0%;  Pred. No. 8.5e-57;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 120

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 299

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
Db        300 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 355

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
Db        356 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 404

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db        405 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 435


RESULT 9
US-09-084-517-23
; Sequence 23, Application US/09084517
; Patent No. 6613329
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICLE DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAVERSTOCK, MEDLEN & CARROLL
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,517
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CARROLL, PETER G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  OPHD-01610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-084-517-23

  Query Match             32.0%;  Score 815;  DB 4;  Length 438;
  Best Local Similarity   40.0%;  Pred. No. 8.5e-57;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |   | :| | |:
Db          4 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 63

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db         64 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 120

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
              |||||      : : |:||   |||||||:|  :||||:|| |: ||||| |
Db        121 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 180

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db        181 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 240

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |    | |:   :   : |    || :    |  :|  :|| |
Db        241 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 299

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :||||| ::    | |||||  | :|:|:  : :: ||:    :   ::  |    | |
Db        300 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 355

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
                |: :| | :  |  |:: | : :    |  | :|           :|
Db        356 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 404

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              | || |:|    :   |||:|:|||||:|| |
Db        405 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 435


RESULT 10
US-08-480-604A-28

; Sequence 28, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-28

  Query Match             32.0%;  Score 815;  DB 1;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 3.7e-56;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |   | :| | |:
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 978

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
```

```
                 |||||  |     : : |:||   |||||||:|  :||||:|| |: |||||  |
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |   | |:   :   : |    || :    |  :|  :|| |
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 1157

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
               :||||| ::    | |||||  | :|:|:  : :: ||:    :   ::  |    | |
Db       1158 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 1213

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
                  |: :| |  :   |  |:: | : :    |  |  :|          :|
Db       1214 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 1262

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
               | || |:|    :   |||:|:|||||:|| |
Db       1263 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 1293


RESULT 11
US-08-405-496A-28
; Sequence 28, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1296 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-28

  Query Match             32.0%;  Score 815;  DB 2;  Length 1296;
  Best Local Similarity   40.0%;  Pred. No. 3.7e-56;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |   | :| | |:
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |        :||:||::||| :||:| ||| |:      | ||||||:|::|:||||||:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 978

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
                 ||||| |     : : |:||   |||||||:|  :||||:|| |: |||||  |
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |   | |:   :   : |    || :    |  :|  :|| |
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 1157

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
               :||||| ::    | |||||  | :|:|:  : :: ||:    :   ::  |    | |
Db       1158 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 1213

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
                  |: :| |  :   |  |:: | : :    |  |  :|          :|
Db       1214 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 1262

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
               | || |:|    :   |||:|:|||||:|| |
Db       1263 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 1293


RESULT 12
US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,136
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,604
;      FILING DATE:
;    PRIOR APPLICATION DATA:
```

; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-28

Query Match 32.0%; Score 815; DB 3; Length 1296;
Best Local Similarity 40.0%; Pred. No. 3.7e-56;
Matches 181; Conservative 77; Mismatches 165; Indels 30; Gaps 11;

```
Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |    | :| | |:
Db        862 LLSTFTEYIKNIIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
                | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 978

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
                 ||||| |     : : |:||   |||||||:|  :||||:|| |: ||||| |
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |   | |:   :   :  |    || :    |  :|  :|| |
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 1157

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
               :||||| ::    | |||||  | :|:|:  : :: ||:    :   ::  |    | |
Db       1158 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 1213

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
                 |: :| |  :  |  |:: | : :    |  |  :|           :|
Db       1214 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 1262

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
               | || |:|    :   |||:|:|||||:|| |
Db       1263 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 1293
```

RESULT 13
US-09-084-517-28
; Sequence 28, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICLE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: OPHD-01610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-517-28

Query Match 32.0%; Score 815; DB 4; Length 1296;
Best Local Similarity 40.0%; Pred. No. 3.7e-56;
Matches 181; Conservative 77; Mismatches 165; Indels 30; Gaps 11;

```
Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |    | :| | |:
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
                | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 978

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
                 ||||| |     : : |:||   |||||||:|  :||||:|| |: ||||| |
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |   | |:   :   :  |    || :    |  :|  :|| |
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 1157
```

```
Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
Db       1158 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 1213

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
Db       1214 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 1262

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db       1263 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 1293


RESULT 14
US-07-618-312A-2
; Sequence 2, Application US/07618312A
; Patent No. 5389540
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  14th Floor
;      STREET:  2200 Clarendon Boulevard,
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/618,312A
;      FILING DATE:  19910516
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crawford Mr, Arthur R
;      REGISTRATION NUMBER:  25,327
;      REFERENCE/DOCKET NUMBER:  510-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  0101 703 8750400
;      TELEFAX:  0101 703 5253468
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-618-312A-2

  Query Match             25.2%;  Score 643.5;  DB 1;  Length 452;
  Best Local Similarity   34.4%;  Pred. No. 3.8e-43;
  Matches  156;  Conservative   76;  Mismatches  179;  Indels   43;  Gaps   14;

Qy         41 ISSNAILSLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKF 99
Db         18 LKKSTILNLDINNDINIDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMD 77

Qy        100 VVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRI 154
Db         78 IEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNL 137

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKIL 213
Db        138 IWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEIT 196

Qy        214 NLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWG 273
Db        197 GLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWG 256

Qy        274 NPLRYDTQYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFII 329
Db        257 NPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFII 314

Qy        330 KKASNSRNINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPIND 384
Db        315 KRYTPNNEI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP--- 369

Qy        385 DPTFYDVLQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNY 438
Db        370 GIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDI 420

Qy        439 FCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 472
Db        421 LIASNWYFNHLKDKI----LGCDWYFVPTDEGWT 450


RESULT 15
US-08-280-228-2
; Sequence 2, Application US/08280228
; Patent No. 5571694
;  GENERAL INFORMATION:
;    APPLICANT:  Makoff Dr, Andrew J
;    APPLICANT:  Romanos Dr, Michael A
;    APPLICANT:  Clare Dr, Jeffrey J
;    APPLICANT:  Fairweather Dr, Neil F
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 No. 5571694th Glebe Road
;      CITY:  Arlington,
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,228
;      FILING DATE:  25-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/618,312
;      FILING DATE:  27-NOV-1990
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8926832.0
;      FILING DATE:  28-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006097.1
;      FILING DATE:  17-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mary J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  117-163
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-280-228-2

  Query Match             25.2%;  Score 643.5;  DB 1;  Length 452;
  Best Local Similarity   34.4%;  Pred. No. 3.8e-43;
  Matches  156;  Conservative   76;  Mismatches  179;  Indels   43;  Gaps   14;

Qy         41 ISSNAILSLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKF 99
              :  : ||:|       || ||: :::    |  :   |       | |:|:| :  |::
Db         18 LKKSTILNLDINNDINIDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMD 77

Qy        100 VVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRI 154
              : |: ||:||:::||:| || : : :: |  |||:||| :|       ||| ||:||| :
Db         78 IEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNL 137

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKIL 213
              |||| |   : : | |   : |  : |: |||  |||||||| :||:|||| |  | :|
Db        138 IWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEIT 196

Qy        214 NLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWG 273
               | |   |:|  ||  | :  ::| |  | || : ||  |:  ||    |   |:||||
Db        197 GLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWG 256

Qy        274 NPLRYDTQYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFII 329
              |||||||:|||       ::: :|  :         ||  :: |  :|  |: || ||:|||
Db        257 NPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFII 314

Qy        330 KKASNSRNINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPIND 384
              |: : :  |  |: |: ||:| | :   ::|    |        |   |    : ||
Db        315 KRYTPNNEI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP--- 369

Qy        385 DPTFYDVLQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNY 438
                |  ::  |  :  ||: |: | :    : || |     ||          |   :
Db        370 GIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDI 420

Qy        439 FCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 472
               | ||   : : |    |||:| |:| |||||
Db        421 LIASNWYFNHLKDKI----LGCDWYFVPTDEGWT 450


Search completed: September 10, 2004, 03:47:17
Job time : 11.0395 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:47:28 ; Search time 91.5581 Seconds
(without alignments)
1656.721 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 2550
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2550	100.0	473	12	US-10-272-898-79	Sequence 79, Appl
2	2550	100.0	473	15	US-10-354-774-79	Sequence 79, Appl
3	2550	100.0	473	15	US-10-271-012-79	Sequence 79, Appl
4	2550	100.0	473	16	US-10-729-122-79	Sequence 79, Appl
5	2550	100.0	473	16	US-10-729-039-79	Sequence 79, Appl
6	2391	93.8	449	10	US-09-910-186A-18	Sequence 18, Appl
7	2387	93.6	1297	12	US-10-272-898-77	Sequence 77, Appl
8	2387	93.6	1297	15	US-10-354-774-77	Sequence 77, Appl
9	2387	93.6	1297	15	US-10-271-012-77	Sequence 77, Appl
10	2387	93.6	1297	15	US-10-452-024-7	Sequence 7, Appli
11	2387	93.6	1297	15	US-10-452-024-150	Sequence 150, App
12	2387	93.6	1297	15	US-10-205-516-14	Sequence 14, Appl
13	2387	93.6	1297	16	US-10-729-122-77	Sequence 77, Appl
14	2387	93.6	1297	16	US-10-729-039-77	Sequence 77, Appl
15	2387	93.6	1307	15	US-10-205-516-28	Sequence 28, Appl
16	2387	93.6	2017	15	US-10-452-024-146	Sequence 146, App
17	1369.5	53.7	472	12	US-10-272-898-46	Sequence 46, Appl
18	1369.5	53.7	472	15	US-10-354-774-46	Sequence 46, Appl
19	1369.5	53.7	472	15	US-10-271-012-46	Sequence 46, Appl
20	1369.5	53.7	472	16	US-10-729-122-46	Sequence 46, Appl
21	1369.5	53.7	472	16	US-10-729-039-46	Sequence 46, Appl
22	1350.5	53.0	472	12	US-10-272-898-44	Sequence 44, Appl
23	1350.5	53.0	472	15	US-10-354-774-44	Sequence 44, Appl
24	1350.5	53.0	472	15	US-10-271-012-44	Sequence 44, Appl
25	1350.5	53.0	472	16	US-10-729-122-44	Sequence 44, Appl
26	1350.5	53.0	472	16	US-10-729-039-44	Sequence 44, Appl
27	1206.5	47.3	1290	15	US-10-360-101-220	Sequence 220, App
28	1206.5	47.3	1290	15	US-10-452-024-119	Sequence 119, App
29	1206.5	47.3	1291	12	US-10-272-898-42	Sequence 42, Appl
30	1206.5	47.3	1291	15	US-10-354-774-42	Sequence 42, Appl
31	1206.5	47.3	1291	15	US-10-271-012-42	Sequence 42, Appl
32	1206.5	47.3	1291	15	US-10-452-024-2	Sequence 2, Appli
33	1206.5	47.3	1291	15	US-10-452-024-118	Sequence 118, App
34	1206.5	47.3	1291	16	US-10-729-122-42	Sequence 42, Appl
35	1206.5	47.3	1291	16	US-10-729-039-42	Sequence 42, Appl
36	1202.5	47.2	1291	15	US-10-452-024-121	Sequence 121, App
37	1201.5	47.1	1291	15	US-10-452-024-123	Sequence 123, App
38	1198.5	47.0	1291	15	US-10-452-024-122	Sequence 122, App
39	1196.5	46.9	1291	15	US-10-452-024-142	Sequence 142, App
40	1177.5	46.2	1291	12	US-10-272-898-40	Sequence 40, Appl
41	1177.5	46.2	1291	15	US-10-354-774-40	Sequence 40, Appl
42	1177.5	46.2	1291	15	US-10-271-012-40	Sequence 40, Appl
43	1177.5	46.2	1291	15	US-10-452-024-117	Sequence 117, App
44	1177.5	46.2	1291	15	US-10-205-516-4	Sequence 4, Appli
45	1177.5	46.2	1291	16	US-10-729-122-40	Sequence 40, Appl

ALIGNMENTS

```
RESULT 1
US-10-272-898-79
; Sequence 79, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 79:
;       SEQUENCE CHARACTERISTICS:
```

```
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-272-898-79

  Query Match             100.0%;  Score 2550;  DB 12;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-194;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 2
US-10-354-774-79
; Sequence 79, Application US/10354774
; Publication No. US20030215468A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/354,774
;             FILING DATE: 30-Jan-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 28-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-354-774-79

  Query Match             100.0%;  Score 2550;  DB 15;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-194;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 3
US-10-271-012-79
; Sequence 79, Application US/10271012
; Publication No. US20030219457A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/271,012
;             FILING DATE: 15-Oct-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-271-012-79

  Query Match             100.0%;  Score 2550;  DB 15;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-194;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 4
US-10-729-122-79
; Sequence 79, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-729-122-79

  Query Match             100.0%;  Score 2550;  DB 16;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-194;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 5
US-10-729-039-79
```

```
; Sequence 79, Application US/10729039
; Publication No. US20040142455A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/729,039
;            FILING DATE: 05-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 1996/08/28
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 79:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 473 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-729-039-79

  Query Match            100.0%;  Score 2550;  DB 16;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-194;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGHHHHHHHHHHSSGHIEGRHMASMADTILIQVFNNYISNISSNAILSLSYRGGRLIDSS 60

Qy         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GYGATMNVGSDVIFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY 120

Qy        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISD 180

Qy        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YINKWFSITITNDRLGNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWI 240

Qy        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKA 300

Qy        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDE 360

Qy        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFG 420

Qy        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473


RESULT 6
US-09-910-186A-18
; Sequence 18, Application US/09910186A
; Publication No. US20030009025A1
; GENERAL INFORMATION:
;  APPLICANT: U.S. Army Medical Research & Material Command
;  TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;  TITLE OF INVENTION:  NEUROTOXIN
;  FILE REFERENCE: A33626-A  067252.0107
;  CURRENT APPLICATION NUMBER: US/09/910,186A
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/US00/12890
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/611,419
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: 60/133,865
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,866
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,867
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,868
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,869
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 60/133,873
;  PRIOR FILING DATE: 1999-05-12
;  PRIOR APPLICATION NUMBER: 08/123,975
;  PRIOR FILING DATE: 1993-09-21
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 449
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-09-910-186A-18

  Query Match            93.8%;  Score 2391;  DB 10;  Length 449;
  Best Local Similarity  99.8%;  Pred. No. 6.7e-182;
  Matches  448;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         25 MADTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 84
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKL 60

Qy         85 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 120

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 180

Qy        205 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSS 240

Qy        265 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLG 300

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 360
```

```
Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 420

Qy        445 YLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||||
Db        421 YLRRISENINKLRLGCNWQFIPVDEGWTE 449


RESULT 7
US-10-272-898-77
; Sequence 77, Application US/10272898
; Publication No. US20030162745A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/272,898
;            FILING DATE: 04-Dec-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 77:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1297 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-272-898-77

  Query Match             93.6%;  Score 2387;  DB 12;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-181;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 8
US-10-354-774-77
; Sequence 77, Application US/10354774
; Publication No. US20030215468A1
;  GENERAL INFORMATION:
;       APPLICANT: Williams, James A.
;                  Thalley, Bruce S.
;       TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                           Botulinum Neurotoxin
;       NUMBER OF SEQUENCES: 82
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Medlen & Carroll
;            STREET: 220 Montgomery Street, Suite 2200
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: United States of America
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/354,774
;            FILING DATE: 30-Jan-2003
;            CLASSIFICATION: 424
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/704,159
;            FILING DATE: 28-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ingolia, Diane E.
;            REGISTRATION NUMBER: 40,027
;            REFERENCE/DOCKET NUMBER: OPHD-02304
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 705-8410
;            TELEFAX: (415) 397-8338
;  INFORMATION FOR SEQ ID NO: 77:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1297 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-354-774-77

  Query Match             93.6%;  Score 2387;  DB 15;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-181;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
                 ||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 9
US-10-271-012-77
; Sequence 77, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-271-012-77

```
  Query Match           93.6%;  Score 2387;  DB 15;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 5.4e-181;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
               ||||||||||||||||||||||||||
Db        1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 10
US-10-452-024-7
; Sequence 7, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1297
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-7

```
  Query Match           93.6%;  Score 2387;  DB 15;  Length 1297;
  Best Local Similarity  100.0%;  Pred. No. 5.4e-181;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 11
US-10-452-024-150
; Sequence 150, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowych, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 150
; LENGTH: 1297
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-452-024-150

Query Match 93.6%; Score 2387; DB 15; Length 1297;
Best Local Similarity 100.0%; Pred. No. 5.4e-181;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 12
US-10-205-516-14
; Sequence 14, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1297
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (7)
; OTHER INFORMATION: mixed sequence
US-10-205-516-14

Query Match 93.6%; Score 2387; DB 15; Length 1297;
Best Local Similarity 100.0%; Pred. No. 5.4e-181;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
```

```
                    ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 13
US-10-729-122-77
; Sequence 77, Application US/10729122
; Publication No. US20040115215A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,122
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-122-77

  Query Match             93.6%;  Score 2387;  DB 16;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-181;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
               ||||||||||||||||||||||||||||
Db        1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 14
US-10-729-039-77
; Sequence 77, Application US/10729039
; Publication No. US20040142455A1
;   GENERAL INFORMATION:
;        APPLICANT: Williams, James A.
;                   Thalley, Bruce S.
;        TITLE OF INVENTION: Multivalent Vaccine For Clostridium
;                            Botulinum Neurotoxin
;        NUMBER OF SEQUENCES: 82
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/729,039
;             FILING DATE: 05-Dec-2003
;             CLASSIFICATION: 424
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/704,159
;             FILING DATE: 1996/08/28
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: OPHD-02304
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 77:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1297 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-729-039-77

  Query Match             93.6%;  Score 2387;  DB 16;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-181;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030
```

```
Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297
```

RESULT 15
US-10-205-516-28
; Sequence 28, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: JZbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1307
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (7)
; OTHER INFORMATION: mixed sequence
US-10-205-516-28

Query Match 93.6%; Score 2387; DB 15; Length 1307;
Best Local Similarity 100.0%; Pred. No. 5.5e-181;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 914

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 974

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1034

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1035 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1094

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1154

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1214

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1274

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              ||||||||||||||||||||||||||||
Db       1275 RRISENINKLRLGCNWQFIPVDEGWTE 1301
```

Search completed: September 10, 2004, 04:58:24
Job time : 93.5581 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:36:11 ; Search time 8.35625 Seconds
(without alignments)
5444.856 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 2550
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2387	93.6	1297	2	S39791	neurotoxin - Clost
2	1206.5	47.3	1291	1	A48940	bontoxilysin (EC 3
3	1177.5	46.2	1291	2	I40631	non-proteolytic bo
4	815	32.0	1296	1	BTCLAB	bontoxilysin (EC 3
5	787	30.9	1296	2	I40645	botulinum neurotox
6	747	29.3	1252	2	S21178	botulinum neurotox
7	744.5	29.2	1268	2	S33411	botulinum neurotox
8	735.5	28.8	1251	2	JH0256	botulinum neurotox
9	668.5	26.2	1274	2	I40813	neurotoxin type F
10	642.5	25.2	1315	1	BTCLTN	tentoxilysin (EC 3
11	612	24.0	1276	2	S11455	botulinum neurotox
12	575.5	22.6	1291	2	S46431	botulinum neurotox
13	575.5	22.6	1291	2	A49777	botulinum neurotox
14	543.5	21.3	1285	2	S70582	botulinum neurotox
15	321	12.6	367	2	S48106	neurotoxin type E
16	320	12.5	366	2	S48110	neurotoxin type F
17	282.5	11.1	369	2	S48109	neurotoxin type F
18	224	8.8	1162	2	I40817	botulinum toxin no
19	212	8.3	1162	2	A47708	progenitor toxin n
20	197	7.7	1196	2	JQ1467	toxin, nontoxic co
21	197	7.7	1196	2	S46430	botulinum neurotox
22	192.5	7.5	1193	2	S68218	botulinum neurotox
23	173.5	6.8	763	2	S17998	gene COX1 intron 4
24	172.5	6.8	1193	2	JC4901	nontoxic-nonhemagg
25	172	6.7	1165	2	I40644	botulinum neurotox
26	164	6.4	588	2	A71661	hypothetical prote
27	161.5	6.3	4688	2	F82885	hypothetical prote
28	158	6.2	960	2	S72284	DNA-directed RNA p
29	158	6.2	2710	2	A37052	toxin A - Clostrid
30	157	6.2	1817	2	H71611	probable secreted
31	154	6.0	3973	2	B71612	hypothetical prote
32	152	6.0	2136	2	A05037	hypothetical prote
33	150.5	5.9	2529	2	B64635	toxin-like outer m
34	149.5	5.9	621	2	E82912	hypothetical prote
35	148	5.8	1127	2	T28317	ORF MSV156 hypothe
36	147	5.8	434	2	H90131	regulator of epide
37	145.5	5.7	377	2	H82882	hypothetical prote
38	144	5.6	760	2	T28224	ORF MSV063 probabl
39	144	5.6	870	2	S27514	mosquitocidal toxi
40	143.5	5.6	2399	2	H71879	toxin-like outer m
41	142	5.6	747	2	H82943	hypothetical prote
42	142	5.6	1802	2	G71616	hypothetical prote
43	141	5.5	495	2	E70198	hypothetical prote
44	141	5.5	2206	2	G71611	hypothetical prote
45	140.5	5.5	3394	2	T18501	hypothetical prote

ALIGNMENTS

RESULT 1
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 93.6%; Score 2387; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 8.7e-131;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         851 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 910

Qy          87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         911 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 970

Qy         147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         971 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1030

Qy         207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1031 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1090

Qy         267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1091 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1150

Qy         327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1151 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 1210

Qy         387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1211 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1270

Qy         447 RRISENINKLRLGCNWQFIPVDEGWTE 473
               ||||||||||||||||||||||||||||
```

```
Db       1271 RRISENINKLRLGCNWQFIPVDEGWTE 1297


RESULT 2
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match           47.3%;  Score 1206.5;  DB 1;  Length 1291;
  Best Local Similarity  50.7%;  Pred. No. 2.5e-62;
  Matches  229;  Conservative   95;  Mismatches  117;  Indels   11;  Gaps    7;

Qy        27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
             |||||::|| | | | :| ||:| |:   ||| ||||| : |   | ||    |||| :
Db       846 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 903

Qy        87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
             | || |   |:: ::::|:| :|||::||:| ||| |: || |: |||||:|:||:|||| 
Db       904 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy       147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
             :||:||||||||||:| |:|||:||||:|:::||:|||:|| :||||: | || ||||| |
Db       964 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 1022

Qy       207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
             : :  | ::  : :: :| |||    | |:|:|:| |:|| ||: : :   | ||| :
Db      1023 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 1082

Qy       267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
             |||||||| |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db      1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 1142

Qy       325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :||||||:|
Db      1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 1200

Qy       385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
               ||: :|||:| |: ||:||:| :||    |   || || :| :  | |:: | :||||
Db      1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 1259

Qy       442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
             |:|||: :      |:|||||||| ||||||
Db      1260 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1291


RESULT 3
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinu
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
```

```
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalytic s
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

  Query Match             46.2%;  Score 1177.5;  DB 2;  Length 1291;
  Best Local Similarity   50.0%;  Pred. No. 1.2e-60;
  Matches  225;  Conservative   94;  Mismatches  120;  Indels   11;  Gaps    7;

Qy         29 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSE 88
              |||::|| | | | :| ||:| ||   ||| ||||| : |   |  ||    |||| :|
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         89 NSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVS 148
              :| |   |:: ::::||| :||::||:| ||| |:||| |: |||||||:|:||:||||:|
Db        906 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        149 IKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIYNGSLKK 208
              |:|||||||||:| |:||:||||:|:::||:|||:|| :||||: | || |||||:|:
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYIYNGTLES 1024

Qy        209 SEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTL 268
              :  | ::  :  : :| |||    | |:|:|:| |:||  :|| : :  :| ||| :  |
Db       1025 NMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYL 1084

Qy        269 KDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLR 326
              |||||||| |: :||:|| | :| |||    :|:||   |: :|  :  |||:|||:| :
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEK 1144

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
              |||:: |||::| ||:|||: |||:|::  :  | :|||     || : :|||: |:|
Db       1145 FIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQSEEKLFLSIISDSN 1202

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 443
               ||  ::||:| |: :|:||:| :||   |   || || :| :  | :   | :|||||:
Db       1203 EFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GVLRKKYKDYFCISK 1261

Qy        444 WYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |||: :     |  |||||||| ||||||
Db       1262 WYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 4
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxi
A;Reference number: S09492; MUID:90235864; PMID:2185020
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin component
A;Reference number: S67988; MUID:96096783; PMID:8521962
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-ter
A;Reference number: A33401; MUID:89350959; PMID:2669749
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, a
A;Reference number: A53884; MUID:94000342; PMID:8397793
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after e
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and aro
A;Reference number: A60025; MUID:91120847; PMID:2126206
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495; PMID:8294407
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

  Query Match             32.0%;  Score 815;  DB 1;  Length 1296;
```

```
  Best Local Similarity   40.0%;  Pred. No. 1.3e-39;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |   | :| | |:
Db        862 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 921

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db        922 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 978

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
              |||||:|     : : |:||   |||||||:|  :||||:|| |: |||||:|
Db        979 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1038

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::||:| |||  | || :::||| ||:| :|||  |:  ||  ||::  ||
Db       1039 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1098

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324
              ||||: |:||  ||: |    | |:   :   :  |    || :    |  :|  :|| |
Db       1099 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 1157

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :||||| ::    | |||||  | :|:|:  : :: ||:    :   ::  |    | |
Db       1158 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 1213

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440
                 |: :| |  :  |  |:: | : :    |  |  :|           :|
Db       1214 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 1262

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              | || |:|    :   |||:|:|||||:|| |
Db       1263 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 1293


RESULT 5
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603; PMID:8310180
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             30.9%;  Score 787;  DB 2;  Length 1296;
  Best Local Similarity   38.1%;  Pred. No. 5.3e-38;
  Matches  174;  Conservative   77;  Mismatches  168;  Indels   38;  Gaps   10;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :|||: |:    ||| | ||| :|:|  | :: |   | || | |:
Db        862 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 921

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:|::||| :||:: ||| :   :  | ||||||:||:|:||||||:
Db        922 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSL 978

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209
              |||||  |     : : |:||   |||||||:|  :||||:||  : ||||| |
Db        979 NYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQ 1038

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269
              : | ||  |::|| | |||  | |  ::: || ||:| :|||  |:  ||  ||::  ||
Db       1039 KPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1098

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIK---------YFSKASMGETAPRTNFNNAAINYQN 320
              |||||  |:||  ||: |  |  | |:           : |    |       : |:
Db       1099 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNS-----T 1153

Qy        321 LYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLA 380
              || | :||||| ::    | |||||  | :|:|:  : :: ||:    :   ::  |
Db       1154 LYEGTKFIIKKYASG---NEDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSAL 1209

Qy        381 PINDDPTFYDVLQIKKYYEKTTYN-CQILCEKDTKTFGLFGIGKFVKDYGYV-WDTYDNY 438
              | |       |: :|   ::   | |::  : :               | |:: :  |||
Db       1210 EIPDVGNLSQVVVMKSKDDQGIRNKCKMNLQDNNG----------NDIGFIGFHLYDNI 1258

Qy        439 --FCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
                 | || |:: :      ||:|:|||||:|| |
Db       1259 AKLVASNWYNRQVGK--ASRTFGCSWEFIPVDDGWGE 1293


RESULT 6
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922; PMID:1541280
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400; PMID:2160960
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918; PMID:2116911
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
```

```
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

  Query Match            29.3%;  Score 747;  DB 2;  Length 1252;
  Best Local Similarity  36.3%;  Pred. No. 1.1e-35;
  Matches  168;  Conservative   80;  Mismatches  155;  Indels   60;  Gaps   14;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  || :   | |:::|:: |:  : :|:||| : :|:  ||          || : |
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
               : | :   |: :::||: : ||||:|||| | |:|  :   : |||||:|:: |:|||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |||:  | ||||| |    :: : | |   : |||||||  :|||||||:: :||||:
Db        951 KVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : ||||  |: |::| ||::||:  |::: |: |||| :||: ||: :||  : :|
Db       1011 LIDQKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325
              | ||||||| | || :||| |    | :|          |    |  :  :    || |:
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFI----DRRKDSTLSINNIRSTILLANRLYSGI 1125

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 385
              :  |:: :||    |||:||: | :|:|          | ||   | ||
Db       1126 KVKIQRVNNSS--TNDNLVRKNDQVYIN-------------FVASK---THLF------- 1160

Qy        386 PTFYDVLQIKKYYEKT--------TYNCQIL-------CEKDTKTFGLFGIGKFVKDYGY 430
              | : |     |  |||          :|  ::      |  : |      ||      |:
Db       1161 PLYADTATTNK--EKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLL----GF 1214

Qy        431 VWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
                ||       | ||   : :: |    || | ||  : || |
Db       1215 KADT----VVASTWYYTHMRDHTNS--NGCFWNFISEEHGWQE 1251


RESULT 7
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match            29.2%;  Score 744.5;  DB 2;  Length 1268;
  Best Local Similarity  34.1%;  Pred. No. 1.5e-35;
  Matches  162;  Conservative   85;  Mismatches  151;  Indels   77;  Gaps   12;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  :   :   |  ::||:: |   | |||||||: ::::  |:          ||:::
Db        843 DKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYS 902

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
              |  | :   |:  ::|:| : |||::|||| |||||       | ||||||:|:: |:|||
Db        903 SRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN---LKNLNNEYTIINCMRNNNSGW 959

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |:|:  | ||||| |    :: : | |:   :|||||||  :||||:|||:: :||||:
Db        960 KISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGN 1019

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : ||||  |: |::| ||::||:  |::: |: |||| :||: ||: :||  : :|
Db       1020 LTDQKSILNLGNIHVDDNILFKIVGCND-TRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIK------------YFSKASMGETAPRTNFN 312
              |||||||  | |: :||| |    |: :                :|| ::        |:
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNI--------FS 1130

Qy        313 NAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKE 372
              ||       || |:  ||:|  ::   | || ||: | :|:|: : : | |::|  |::
Db       1131 NA-----RLYTGVEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSE-YQLYADVSTSA 1184

Qy        373 IQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQIL---------CEKDTKT-----FGL 418
              ::  : |  |::                : ||   :         |  : ||      ||
Db       1185 VEKTIKLRRISN---------------SNYNSNQMIIMDSIGDNCTMNFKTNNGNDIGL 1228

Qy        419 FGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |               : |    | || : |  |       || | ||  : || |
Db       1229 LGF------------HLNNLVASSWYYKNIRNNTR--NNGCFWSFISKEHGWQE 1268


RESULT 8
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum typ
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E tox
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the li
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

  Query Match            28.8%;  Score 735.5;  DB 2;  Length 1251;
  Best Local Similarity  36.0%;  Pred. No. 4.9e-35;
  Matches  162;  Conservative   81;  Mismatches  172;  Indels   35;  Gaps   10;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  || :   | |:::|:: |:  : :|:|||| : :|:  ||          || : |
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
               : | :   |: :::||: : ||||:|||| | |:|  :   : |||||:|:: |:|||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |||:  | ||||| |  :  :: : | |   : |||||||  :|||||||:: :||||:
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010
```

```
Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
Db       1011 LIDKKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFI----NRRTDSTLSINNIRSTILLANRLYSGI 1125

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLF--LAPIN 383
Db       1126 KVKIQRVNNSS--TNDNLVRKNDQVYIN------------FVASKTHLLPLYADTATTN 1170

Qy        384 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 443
Db       1171 KEKTIKISSSGNRFNQVVVMNSVGNCTMNFKNNNGNNIGLL----GFKADT----VVAST 1222

Qy        444 WYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db       1223 WYYTHMRDNTNS--NGFFWNFISEEHGWQE 1250
```

RESULT 9
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match   26.2%; Score 668.5; DB 2; Length 1274;
  Best Local Similarity   34.3%; Pred. No. 3.9e-31;
  Matches 158; Conservative   78; Mismatches 175; Indels   49; Gaps   13;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
Db        851 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KN 141
Db        911 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN-------REYTIINCMGNN 963

Qy        142 DSGWKVSIKGNR---IIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNA 198
Db        964 NSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNS 1023

Qy        199 NIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSL 258
Db       1024 RIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETL 1082

Qy        259 YWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-----KASMGETAPRTNFNN 313
Db       1083 YSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTE-----GS 1137

Qy        314 AAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 373
Db       1138 VFLNYK-LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVE-YRLYADTKSEK- 1194

Qy        374 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWD 433
Db       1195 --EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQ---NNNGSNIGLLGF------------ 1237

Qy        434 TYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db       1238 -HSNNLVASSWYYNNIRRNTSS--NGCFWSSISKENGWKE 1274
```

RESULT 10
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.

C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglic
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 25.2%; Score 642.5; DB 1; Length 1315;
Best Local Similarity 34.1%; Pred. No. 1.3e-29;
Matches 155; Conservative 77; Mismatches 179; Indels 43; Gaps 14;

```
Qy         41 ISSNAILSLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKF 99
Db        881 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMD 940

Qy        100 VVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRI 154
Db        941 IEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNL 1000

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKIL 213
Db       1001 IWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEIT 1059

Qy        214 NLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWG 273
Db       1060 GLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWG 1119

Qy        274 NPLRYDTQYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFII 329
Db       1120 NPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFII 1177

Qy        330 KKASNSRNINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPIND 384
Db       1178 KRYTPNNEI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP--- 1232

Qy        385 DPTFYDVLQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNY 438
Db       1233 GIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDI 1283

Qy        439 FCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 472
Db       1284 LIASNWYFNHLKDKI----LGCDWYFVPTDEGWT 1313
```

RESULT 11
S11455
botulinum neurotoxin type D - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11455
R;Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegls
Nucleic Acids Res. 18, 5556, 1990
A;Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A;Reference number: S11455; MUID:91016853; PMID:2216736
A;Accession: S11455
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-1276 <BIN>
A;Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 24.0%; Score 612; DB 2; Length 1276;
Best Local Similarity 31.0%; Pred. No. 7.3e-28;
Matches 146; Conservative 81; Mismatches 176; Indels 68; Gaps 11;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
Db        910 S-GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINY---QNLYL 323
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI----------APESNV-LVLVQYPDRSKLYT 1131

Qy        324 GLRFIIKKASN----SRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFL 379
Db       1132 GNPITIKSVSDKNPYSRILNGDNII----------LHMLYNSRKYMI---- 1168

Qy        380 APINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY- 438
Db       1169 --IRDTDTIY-ATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFR 1225

Qy        439 --------------FCISQWYLRRISENINKLRLGCN--WQFIPVDEGWTE 473
Db       1226 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276
```

RESULT 12
S46431
botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)
N;Alternate names: BoNT/C1 protein
C;Species: Clostridium botulinum phage 1C
A;Variety: strain C 468
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 16-Jul-1999
C;Accession: S46431; S49107
R;Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.
Mol. Gen. Genet. 243, 631-640, 1994
A;Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic
A;Reference number: S46426; MUID:94301293; PMID:8028579
A;Accession: S46431
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1291 <HAU>
A;Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175
A;Experimental source: strain C 468
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Superfamily: tetanus toxin

Query Match 22.6%; Score 575.5; DB 2; Length 1291;
Best Local Similarity 31.7%; Pred. No. 9.7e-26;
Matches 151; Conservative 84; Mismatches 169; Indels 73; Gaps 16;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
Db        854 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 913

Qy         87 S--ENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144
Db        914 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSG 967

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204
Db        968 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1026

Qy        205 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKF--------VWIKDFNIFGRELNATEVS 256
```

```
Db       1027 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 1086

Qy        257 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAA 315

Db       1087 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE-- 1144

Qy        316 INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI-----------------DNI 357

Db       1145 --------GYKIIIKRI---RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 1193

Qy        358 SDES-YRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTK 414

Db       1194 STEDIYAIGLREQTKDINDNIIFQIQPMNN--TYYYASQI----FKSNFNGENIS----- 1242

Qy        415 TFGLFGIGKFVKDYGYVWDTYDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 469

Db       1243 --GICSIGTYRFRLGGDWYRH-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1291


RESULT 13
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C;Species: Clostridium botulinum
C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 17-Mar-2003
C;Accession: S11291; A35396; S22166; A49777
R;Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A;Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A;Reference number: S11291; MUID:90370487; PMID:2204031
A;Accession: S11291
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-84,'P',86-1291 <HAU>
A;Cross-references: EMBL:X53751; NID:g14905; PIDN:CAA37780.1; PID:g14906
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A;Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxi
A;Reference number: A35396; MUID:91024998; PMID:2222445
A;Accession: A35396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-669,'R',671-1291 <TS1>
R;Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A;Description: Nucleotide sequence of the gene for one of the components of hemagglutini
A;Reference number: S22163
A;Accession: S22166
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <TS2>
A;Cross-references: EMBL:X62389; NID:g558175; PIDN:CAA44263.1; PID:g40390
R;Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A;Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A;Reference number: A49777; MUID:91282468; PMID:2059039
A;Accession: A49777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-607 <TS3>
A;Cross-references: GB:D90210
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             22.6%;  Score 575.5;  DB 2;  Length 1291;
  Best Local Similarity   31.7%;  Pred. No. 9.7e-26;
  Matches  151;  Conservative   84;  Mismatches  169;  Indels   73;  Gaps   16;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Db        854 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 913

Qy         87 S--ENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144

Db        914 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSG 967

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204

Db        968 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1026

Qy        205 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKF--------VWIKDFNIFGRELNATEVS 256

Db       1027 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 1086

Qy        257 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAA 315

Db       1087 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE-- 1144

Qy        316 INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI-----------------DNI 357

Db       1145 --------GYKIIIKRI---RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 1193

Qy        358 SDES-YRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTK 414

Db       1194 STEDIYAIGLREQTKDINDNIIFQIQPMNN--TYYYASQI----FKSNFNGENIS----- 1242

Qy        415 TFGLFGIGKFVKDYGYVWDTYDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 469

Db       1243 --GICSIGTYRFRLGGDWYRH-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1291


RESULT 14
S70582
botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sA
C;Species: Clostridium botulinum phage d-sA
A;Note: host Clostridium botulinum type D (strain South Africa)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70582
R;Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.
Biochim. Biophys. Acta 1307, 123-126, 1996
A;Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C an
A;Reference number: S70582; MUID:96283801; PMID:8679691
A;Accession: S70582
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1285 <MOR>
A;Cross-references: EMBL:D38442; NID:g1374775; PIDN:BAA07477.1; PID:g1374776
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membr
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; neurotoxin; transmembrane protein
F;1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>
F;448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT2>

  Query Match             21.3%;  Score 543.5;  DB 2;  Length 1285;
  Best Local Similarity   31.3%;  Pred. No. 6.9e-24;
  Matches  146;  Conservative   78;  Mismatches  189;  Indels   53;  Gaps   14;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Db        850 NSLLKDMINEYFNSINDSKILSLQNKKNTLMDTSGYNAEVRVEGNVQLNPIFPFDFKLGS 909

Qy         87 S--ENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144

Db        910 SGDDRGKIIVTQNENIVYNAMYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSG 963

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204

Db        964 WSIGIISNFLVFTLKQNENSEQDINFSYDISKNAAGY-NKWFFVTITTNMMGNMMIYING 1022

Qy        205 SLKKSEKILNLDRINSSNDIDFK--------LINCTDTTKFVWIKDFNIFGRELNATEVS 256

Db       1023 KLIDTIKVKELTGINFSKTITFQMNKIPNTGLITSDSDNINMWIRDFYIFAKELDDKDIN 1082

Qy        257 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI 316
```

```
Db       1083 ILFNSLQYTNVVKDYWGNDLRYDKEYYMINVNYMN---RYMSKKGNG-IVFNTRKNNNDF 1138

Qy        317 NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ 376

Db       1139 NE-----GYKIIIKRI---RGNTNDTRVRGENVLYFN-TTIDNKQYSLGMYKPSRNLGTD 1189

Qy        377 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQ----------ILCEKDTKTFGLFGIGKFVK 426

Db       1190 LVPLGALDQP----MDEIRKYGSFIIQPCNTFDYYASQLFLSSNATTNRLGILSIGSYSF 1245

Qy        427 DYGYVWDTYDNYFCISQWYLRRIS-ENINKL--RLGCNWQFIPVDE 469

Db       1246 KLG------DDYWFNHEYLIPVIKIEHYASLLESTSTHWVFVPASE 1285


RESULT 15
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

  Query Match             12.6%;  Score 321;  DB 2;  Length 367;
  Best Local Similarity   35.6%;  Pred. No. 1.2e-11;
  Matches   62;  Conservative   34;  Mismatches   52;  Indels   26;  Gaps    3;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Db        218 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 277

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145

Db        278 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 335

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNAN 199

Db        336 KVSLNHNEIIWTLQDNAGINQKLAFNY----------------------GNAN 366


Search completed: September 10, 2004, 03:44:23
Job time : 10.3563 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:10:35 ; Search time 5.77123 Seconds
(without alignments)
4267.583 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 2550
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2387	93.6	1296	1	BXG_CLOBO	Q60393 clostridium
2	1206.5	47.3	1290	1	BXB_CLOBO	P10844 clostridium
3	815	32.0	1295	1	BXA1_CLOBO	P10845 clostridium
4	787	30.9	1295	1	BXA2_CLOBO	Q45894 clostridium
5	745.5	29.2	1250	1	BXE_CLOBO	Q00496 clostridium
6	735.5	28.8	1250	1	BXE_CLOBU	P30995 clostridium
7	668.5	26.2	1274	1	BXF_CLOBO	P30996 clostridium
8	642.5	25.2	1314	1	TETX_CLOTE	P04958 clostridium
9	612	24.0	1276	1	BXD_CLOBO	P19321 clostridium
10	575.5	22.6	1290	1	BXC1_CLOBO	P18640 clostridium
11	224	8.8	1162	1	BXEN_CLOBU	Q06366 clostridium
12	212	8.3	1162	1	BXEN_CLOBO	P46082 clostridium
13	197	7.7	1196	1	BXCN_CLOBO	P46081 clostridium
14	158	6.2	2710	1	TOXA_CLODI	P16154 clostridium
15	152	6.0	2136	1	YCF2_MARPO	P09975 marchantia
16	144.5	5.7	967	1	Y5G0_CLOAB	P33747 clostridium
17	144	5.6	760	1	ETF2_MSEPV	Q9yw29 melanoplus
18	142	5.6	747	1	Y030_UREPA	Q9prb5 ureaplasma
19	135.5	5.3	861	1	SYL_BUCBP	P59433 buchnera ap
20	135	5.3	850	1	PRSN_PIEBR	Q9gv36 pieris bras
21	135	5.3	2339	1	RPC1_PLAFA	P27625 plasmodium
22	134	5.3	618	1	EX5A_BUCBP	Q89ab2 buchnera ap
23	134	5.3	1179	1	EX5B_BUCAP	Q8k9a9 buchnera ap
24	134	5.3	1310	1	VAC3_HELPY	Q48253 helicobacte
25	133.5	5.2	339	1	RMAR_CANGA	P21358 candida gla
26	132	5.2	337	1	YA76_METJA	Q58476 methanococc
27	132	5.2	735	1	CIGB_DICDI	Q94481 dictyosteli
28	130.5	5.1	617	1	Y237_BUCAI	P57331 buchnera ap
29	130	5.1	917	1	HXA3_HAEIN	P45355 haemophilus
30	127.5	5.0	537	1	ARP_PLAFA	P04931 plasmodium
31	126.5	5.0	850	1	PRSN_PIERA	Q9u8q4 pieris rapa
32	125.5	4.9	1180	1	EX5B_BUCBP	Q89ab3 buchnera ap
33	125	4.9	1956	1	ATX1_PLAFA	Q04956 plasmodium
34	124.5	4.9	528	1	AMY_BACCI	P08137 bacillus ci
35	124	4.9	386	1	RMAR_HANWI	P48849 hansenula w
36	124	4.9	552	1	YC92_METJA	Q58688 methanococc
37	124	4.9	655	1	YKDA_MYCCA	P45615 mycoplasma
38	122.5	4.8	960	1	SYV_BUCAP	Q8k9i1 buchnera ap
39	122.5	4.8	1024	1	RPOB_PLAFA	P21421 plasmodium
40	121.5	4.8	1636	1	YN37_YEAST	P48563 saccharomyc
41	121	4.7	1323	1	ADR1_YEAST	P07248 saccharomyc
42	120	4.7	968	1	RNE_BUCAP	Q8k9j9 buchnera ap
43	119	4.7	368	1	TRMU_WIGBR	Q8d397 wiggleswort
44	119	4.7	987	1	YD94_METJA	Q58789 methanococc
45	118	4.6	473	1	SYE_WIGBR	Q8d375 wiggleswort

ALIGNMENTS

RESULT 1
BXG_CLOBO
ID BXG_CLOBO STANDARD; PRT; 1296 AA.
AC Q60393;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE (Bontoxilysin G).
GN BOTG.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;
RX MEDLINE=94092745; PubMed=8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins.";
RL Biochim. Biophys. Acta 1216:487-491(1993).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H). The light chain has the pharmacological activity,
CC while the N- and C-terminal of the heavy chain mediate channel
CC formation and toxin binding, respectively.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -!- SIMILARITY: Belongs to peptidase family M27.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X74162; CAA52275.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR006025; Pept_M_Zn_BS.

```
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

  Query Match             93.6%;  Score 2387;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-135;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 909

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 969

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 1029

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 1089

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGLR 1149

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1150 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPINDDP 1209

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1210 TFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYL 1269

Qy        447 RRISENINKLRLGCNWQFIPVDEGWTE 473
              |||||||||||||||||||||||||||
Db       1270 RRISENINKLRLGCNWQFIPVDEGWTE 1296


RESULT 2
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; A48940; A48940.
DR   PDB; 1EPW; 01-NOV-00.
DR   PDB; 1F31; 01-NOV-00.
DR   PDB; 1F82; 16-AUG-00.
DR   PDB; 1F83; 16-AUG-00.
DR   PDB; 1FQH; 06-DEC-00.
DR   PDB; 1G9A; 13-NOV-02.
DR   PDB; 1G9B; 13-NOV-02.
DR   PDB; 1G9C; 13-NOV-02.
DR   PDB; 1G9D; 13-NOV-02.
DR   PDB; 1I1E; 21-NOV-01.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;

  Query Match             47.3%;  Score 1206.5;  DB 1;  Length 1290;
  Best Local Similarity   50.7%;  Pred. No. 1e-64;
  Matches  229;  Conservative   95;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              |||||::|| | | | :| ||:| |:   ||| ||||| : |    |  ||    |||| :
Db        845 DTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELND--KNQFKLTS 902

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: ::::|:| :||::||:| ||| |: || |: ||||||:|:||:||||
Db        903 SANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 962

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:||||||||||:| |:||:||||:|:::||:|||:|| :||||: | || ||||| |
Db        963 ISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LNNAKIYINGKL 1021

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              : :  | ::  : :: :| |||     | |:|:|:| |:||  ||: : :   | ||| :
Db       1022 ESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSYSE 1081

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||| |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db       1082 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 1141

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||     |: : :|||||:|
Db       1142 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYTYKYFKKEEEKLFLAPISD 1199

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
                ||: :|||:| |: ||:||:| :||    |   || || :| :  | |:: | :||||
Db       1200 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFEEYKDYFCI 1258

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :      |:|||||||| ||||||
Db       1259 SKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE 1290
```

```
RESULT 3
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P01561; P18639;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Prevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=Hall;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=Type A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
```

```
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   MEDLINE=21556941; PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   -!- DATABASE: NAME=BOTOX product information Web site;
CC       WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC   -!- DATABASE: NAME=Protein Spotlight;
CC       NOTE=Issue 19 of February 2002;
CC       WWW="http://www.expasy.org/spotlight/articles/sptlt019.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASES ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASES ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             32.0%;  Score 815;  DB 1;  Length 1295;
  Best Local Similarity   40.0%;  Pred. No. 2.4e-41;
  Matches  181;  Conservative   77;  Mismatches  165;  Indels   30;  Gaps   11;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89
              |:  |  || || : :||:| |    ||| | | : :|:|| | |: |   | :| | |:
Db        861 LLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLES 920

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149
              | |       :||:||::||| :||:| ||| |:     | ||||||:|::|:||||||:
Db        921 SKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNS---ISLNNEYTIINCMENNSGWKVSL 977

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYIINGSLKKS 209
                  ||||| |     : : |:||   |||||||:|  :||||:|| |: ||||| |
```

```
Db        978 NYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQ 1037

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269

Db       1038 KPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILK 1097

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIKYFS---KASMGETAPRTNF--NNAAINYQNLYLG 324

Db       1098 DFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLN-SSLYRG 1156

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384

Db       1157 TKFIIKKYASG---NKDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSALEIPD 1212

Qy        385 DPTFYDVLQIK-KYYEKTTYNCQI-LCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY--FC 440

Db       1213 VGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQF----------NNIAKLV 1261

Qy        441 ISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473

Db       1262 ASNWYNRQIER--SSRTLGCSWEFIPVDDGWGE 1292
```

```
RESULT 4
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   PIR; I40645; I40645.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;
```

```
  Query Match             30.9%;  Score 787;  DB 1;  Length 1295;
  Best Local Similarity   38.1%;  Pred. No. 1.1e-39;
  Matches  174;  Conservative   77;  Mismatches  168;  Indels   38;  Gaps   10;

Qy         30 LIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSEN 89

Db        861 LLSTFTEYIKNIVNTSILSIVYKKDDLIDLSRYGAKINIGDRVYYDSIDKNQIKLINLES 920

Qy         90 SNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSI 149

Db        921 STIEVILKNAIVYNSMYENFSTSFWIKIPKYFS---KINLNNEYTIINCIENNSGWKVSL 977

Qy        150 KGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKKS 209

Db        978 NYGEIIWTLQDNKQNIQRVVFKYSQMVNISDYINRWIFVTITNNRLTKSKIYINGRLIDQ 1037

Qy        210 EKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLK 269

Db       1038 KPISNLGNIHASNKIMFKLDGCRDPRRYIMIKYFNLFDKELNEKEIKDLYDSQSNSGILK 1097

Qy        270 DFWGNPLRYDTQYYLFNQGMQNIYIK---------YFSKASMGETAPRTNFNNAAINYQN 320

Db       1098 DFWGNYLQYDKPYYMLNLFDPNKYVDVNNIGIRGYMYLKGPRGSVVTTNIYLNS-----T 1152

Qy        321 LYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLA 380

Db       1153 LYEGTKFIIKKYASG---NEDNIVRNNDRVYINV-VVKNKEYRLATNASQAGVEKILSAL 1208

Qy        381 PINDDPTFYDVLQIKKYYEKTTYN-CQILCEKDTKTFGLFGIGKFVKDYGYV-WDTYDNY 438

Db       1209 EIPDVGNLSQVVVMKSKDDQGIRNKCKMNLQDNNG----------NDIGFIGFHLYDNI 1257

Qy        439 --FCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
```

```
Db       1258 AKLVASNWYNRQVGK--ASRTFGCSWEFIPVDDGWGE 1292


RESULT 5
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beluga;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; S08575; S08575.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;

  Query Match             29.2%;  Score 745.5;  DB 1;  Length 1250;
  Best Local Similarity   36.1%;  Pred. No. 3.3e-37;
  Matches  167;  Conservative   80;  Mismatches  156;  Indels   59;  Gaps   14;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  || :   | |:::|:: |:  : :|:||| : :|:  ||       || : |
Db        832 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 891

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
                : | :   |: :::||: : ||||:|||| | |:|  :   : |||||:|:: |:|||
Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 949

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |||:  | ||||  |    :: : | |   : |||||||  :||||||||:: :||||:
Db        950 KVSLNHNEIIWTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
```

```
Db       1010 LIDQKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1068

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325

Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFI----DRRKDSTLSINNIRSTILLANRLYSGI 1124

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 385

Db       1125 KVKIQRVNNSS--TNDNLVRKNDQVYIN------------FVASK---THLF------- 1159

Qy        386 PTFYDVLQIKKYYEKT--------TYNCQIL------CEKDTKTFGLFGIGKFVKDYGYV 431

Db       1160 PLYADTATTNK--EKTIKISSSGNRFNQVVVMNSVGNCTMNFKNNNGNNIGLL----GFK 1213

Qy        432 WDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473

Db       1214 ADT----VVASTWYYTHMRDHTNS--NGCFWNFISEEHGWQE 1249


RESULT 6
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;

  Query Match             28.8%;  Score 735.5;  DB 1;  Length 1250;
  Best Local Similarity   36.0%;  Pred. No. 1.3e-36;
  Matches  162;  Conservative   81;  Mismatches  172;  Indels   35;  Gaps   10;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86

Db        832 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 891

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145

Db        892 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 949

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205

Db        950 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1009

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265

Db       1010 LIDKKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1068

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325

Db       1069 NILKDFWGNYLLYDKEYYLLNVLKPNNFI----NRRTDSTLSINNIRSTILLANRLYSGI 1124

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLF--LAPIN 383

Db       1125 KVKIQRVNNSS--TNDNLVRKNDQVYIN------------FVASKTHLLPLYADTATTN 1169

Qy        384 DDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 443

Db       1170 KEKTIKISSSGNRFNQVVVMNSVGNCTMNFKNNNGNNIGLL----GFKADT----VVAST 1221

Qy        444 WYLRRISENINKLRLGCNWQFIPVDEGWTE 473

Db       1222 WYYTHMRDNTNS--NGFFWNFISEEHGWQE 1249
```

```
RESULT 7
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=Hobbs FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   PIR; I40813; I40813.
DR   PIR; S48109; S48109.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;

  Query Match             26.2%;  Score 668.5;  DB 1;  Length 1274;
  Best Local Similarity   34.3%;  Pred. No. 1.3e-32;
  Matches  158;  Conservative   78;  Mismatches  175;  Indels   49;  Gaps   13;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||   ||      |  ::|| : |   : || ||||: :::  :|        || : |
Db        851 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYN 910

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KN 141
              |  | :   |:  ::|:| : ||||:|||| ||:    |:|        |||||:|:  |
Db        911 SRLSEVNIAQNNDIIYNSRYQNFSISFWVRIPKHYKPMNHN-------REYTIINCMGNN 963

Qy        142 DSGWKVSIKGNR---IIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNA 198
              :||||:|::  |   ||||| | :   ::: | |   : ||:|||||  :||||:||||:
Db        964 NSGWKISLRTVRDCEIIWTLQDTSGNKENLIFRYEELNRISNYINKWIFVTITNNRLGNS 1023

Qy        199 NIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSL 258
              |||||:|   : | ||  |: |::| ||:: | | | :| |: | :|  ||: ||: :|
Db       1024 RIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETL 1082

Qy        259 YWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFS-----KASMGETAPRTNFNN 313
              |  :   : ||::||| | |: :|||||   :: ||   |          | |       :
Db       1083 YSNEPDPSILKNYWGNYLLYNKKYYLFNLLRKDKYITLNSGILNINQQRGVTE-----GS 1137

Qy        314 AAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEI 373
              :||: || |:  ||:|        | || ||: |  |:|: :   | ||:|    |::
Db       1138 VFLNYK-LYEGVEVIIRKNGPIDISNTDNFVRKNDLAYINVVDRGVE-YRLYADTKSEK- 1194

Qy        374 QTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWD 433
                :   | : :   :: :       | | |     :    || |
Db       1195 --EKIIRTSNLNDSLGQIIVMDSIGNNCTMNFQ---NNNGSNIGLLGF------------ 1237

Qy        434 TYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
               : |    | ||   |  | :    || |  |  : || |
Db       1238 -HSNNLVASSWYYNNIRRNTSS--NGCFWSSISKENGWKE 1274


RESULT 8
TETX_CLOTE
ID   TETX_CLOTE     STANDARD;      PRT;  1314 AA.
AC   P04958;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE   Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
```

```
DE   chain (Tetanus toxin chain H)].
GN   TETX OR CTP60.
OS   Clostridium tetani.
OG   Plasmid pE88, and Plasmid 75 Kbp.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=87053814; PubMed=3536478;
RA   Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA   Weller U., Hudel M., Habermann E., Niemann H.;
RT   "Tetanus toxin: primary structure, expression in E. coli, and
RT   homology with botulinum toxins.";
RL   EMBO J. 5:2495-2502(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN3911; PLASMID=75 Kbp;
RX   MEDLINE=87040747; PubMed=3774547;
RA   Fairweather N.F., Lyness V.A.;
RT   "The complete nucleotide sequence of tetanus toxin.";
RL   Nucleic Acids Res. 14:7809-7812(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88; PLASMID=pE88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN   [4]
RP   SEQUENCE OF 742-1314 FROM N.A.
RC   PLASMID=75 Kbp;
RX   MEDLINE=86085672; PubMed=3510187;
RA   Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT   "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT   fragment C in Escherichia coli.";
RL   J. Bacteriol. 165:21-27(1986).
RN   [5]
RP   PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX   MEDLINE=90201034; PubMed=2108021;
RA   Krieglstein K., Henschen A., Weller U., Habermann E.;
RT   "Arrangement of disulfide bridges and positions of sulfhydryl groups
RT   in tetanus toxin.";
RL   Eur. J. Biochem. 188:39-45(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92037649; PubMed=1935979;
RA   Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;
RT   "Limited proteolysis of tetanus toxin. Relation to activity and
RT   identification of cleavage sites.";
RL   Eur. J. Biochem. 202:41-51(1991).
RN   [7]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93010948; PubMed=1396558;
RA   Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA   Montecucco C.;
RT   "Tetanus toxin is a zinc protein and its inhibition of
RT   neurotransmitter release and protease activity depend on zinc.";
RL   EMBO J. 11:3577-3583(1992).
RN   [8]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX   MEDLINE=97475217; PubMed=9334741;
RA   Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA   Sax M.;
RT   "Structure of the receptor binding fragment HC of tetanus
RT   neurotoxin.";
RL   Nat. Struct. Biol. 4:788-792(1997).
CC   -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC       BOND OF SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC       synaptobrevin 2.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC       YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC       AND ARE NON-TOXIC AFTER SEPARATION.
CC   -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC       GANGLIOSIDE RECEPTORS.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04436; CAA28033.1; -.
DR   EMBL; X06214; CAA29564.1; -.
DR   EMBL; AF528097; AAO37454.1; -.
DR   EMBL; M12739; AAA23282.1; -.
DR   PIR; A25689; BTCLTN.
DR   PDB; 1AF9; 29-APR-98.
DR   PDB; 1A8D; 14-OCT-98.
DR   PDB; 1D0H; 27-MAR-00.
DR   PDB; 1DFQ; 24-MAR-00.
DR   PDB; 1DIW; 24-MAR-00.
DR   PDB; 1DLL; 24-MAR-00.
DR   PDB; 1FV3; 05-SEP-01.
DR   MEROPS; M27.001; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW   3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   CHAIN         1    456       TETANUS TOXIN LIGHT CHAIN.
FT   CHAIN       457   1314       TETANUS TOXIN HEAVY CHAIN.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    233    233       BY SIMILARITY.
FT   METAL       236    236       ZINC (CATALYTIC) (BY SIMILARITY).
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    669    689       POTENTIAL.
FT   DISULFID    438    466       INTERCHAIN.
FT   DISULFID   1076   1092
FT   HELIX       876    882
FT   TURN        883    883
FT   STRAND      884    891
FT   TURN        892    893
FT   STRAND      894    897
FT   STRAND      904    907
FT   TURN        909    910
FT   STRAND      912    915
FT   STRAND      920    925
```

```
FT   TURN        928    929
FT   STRAND      932    935
FT   HELIX       938    940
FT   TURN        941    946
FT   STRAND      949    956
FT   HELIX       962    968
FT   TURN        969    970
FT   STRAND      972    977
FT   STRAND      980    981
FT   HELIX       983    985
FT   STRAND      987    995
FT   TURN        996    997
FT   STRAND      998   1004
FT   TURN       1006   1007
FT   STRAND     1010   1016
FT   STRAND     1020   1020
FT   TURN       1021   1022
FT   STRAND     1031   1037
FT   TURN       1039   1040
FT   STRAND     1042   1047
FT   TURN       1048   1049
FT   STRAND     1050   1056
FT   TURN       1058   1059
FT   STRAND     1068   1074
FT   TURN       1079   1080
FT   STRAND     1082   1091
FT   HELIX      1097   1105
FT   TURN       1106   1107
FT   STRAND     1112   1112
FT   STRAND     1114   1114
FT   TURN       1116   1117
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   TURN       1123   1124
FT   STRAND     1127   1131
FT   HELIX      1132   1134
FT   TURN       1135   1136
FT   STRAND     1137   1141
FT   TURN       1144   1145
FT   STRAND     1148   1152
FT   STRAND     1155   1158
FT   TURN       1159   1162
FT   STRAND     1163   1166
FT   STRAND     1173   1178
FT   TURN       1184   1185
FT   STRAND     1188   1188
FT   STRAND     1190   1190

  Query Match           25.2%;  Score 642.5;  DB 1;  Length 1314;
  Best Local Similarity   34.1%;  Pred. No. 4.9e-31;
  Matches  155;  Conservative   77;  Mismatches  179;  Indels   43;  Gaps   14;

Qy         41 ISSNAILSLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKF 99
              :  : ||:|      : | ||: :::    |  :   |      | |:|:| :  |::
Db        880 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMD 939

Qy        100 VVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRI 154
              : |: ||:||:::||:| || : : :: |  |||:||| :|       ||| ||:||| :
Db        940 IEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNL 999

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKIL 213
              |||| |   : : | |   : |  : |: |||  |||||||| :||:|||| |   | :|
Db       1000 IWTLKDSAGEVRQITFR-DLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEIT 1058

Qy        214 NLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWG 273
               |  |   |:|  ||  | :  ::| |  | || : ||  |:  ||    |   |:||||
Db       1059 GLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWG 1118

Qy        274 NPLRYDTQYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFII 329
              |||||||:|||       ::: :|  :       ||  :: |  :|  |: || ||:|||
Db       1119 NPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYNGLKFII 1176

Qy        330 KKASNSRNINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPIND 384
              |: : :  |  |: |: ||:| | :   ::|    |         |   |    : ||
Db       1177 KRYTPNNEI--DSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP--- 1231

Qy        385 DPTFYDVLQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNY 438
                  |  ::  |  :  ||: |: | :    : || |     ||          |   :
Db       1232 GIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGN---------DPNRDI 1282

Qy        439 FCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 472
                | ||   : : |    |||:| |:| |||||
Db       1283 LIASNWYFNHLKDKI----LGCDWYFVPTDEGWT 1312


RESULT 9
BXD_CLOBO
ID   BXD_CLOBO      STANDARD;      PRT;  1276 AA.
AC   P19321;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BoNT/D)
DE   (Bontoxilysin D).
GN   BOTD.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BVD/-3;
RX   MEDLINE=91016853; PubMed=2216736;
RA   Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,
RA   Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT   "Nucleotide sequence of the gene encoding Clostridium botulinum
RT   neurotoxin type D.";
RL   Nucleic Acids Res. 18:5556-5556(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB16;
RX   MEDLINE=93042276; PubMed=1420572;
RA   Sunagawa H., Ohyama T., Watanabe T., Inoue K.;
RT   "The complete amino acid sequence of the Clostridium botulinum type D
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   phage d-16 phi genome.";
RL   J. Vet. Med. Sci. 54:905-913(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   STRAIN=D-SA, and D-1873;
RX   MEDLINE=89339741; PubMed=2668193;
RA   Moriishi K., Syuto B., Kubo S., Oguma K.;
RT   "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT   strains.";
RL   Infect. Immun. 57:2886-2891(1989).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC       SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
```

```
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54254; CAA38175.1; -.
DR   EMBL; S49407; AAB24244.1; -.
DR   PIR; S11455; S11455.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Metalloprotease; Zinc.
FT   CHAIN         1    442       BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT   CHAIN       443   1276       BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    437    450       INTERCHAIN (PROBABLE).
FT   VARIANT      15     16       ND -> PV (IN STRAIN D-SA).
FT   VARIANT      17     18       ND -> LQ (IN STRAIN D-1873).
FT   VARIANT     452    452       K -> Q (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> T (IN STRAIN D-SA).
FT   VARIANT     457    457       R -> F (IN STRAIN D-1873).
FT   VARIANT     462    462       A -> D (IN STRAIN D-1873).
FT   VARIANT     489    489       K -> N (IN STRAIN CB16).
FT   VARIANT     644    644       N -> K (IN STRAIN CB16).
FT   VARIANT    1122   1122       Q -> R (IN STRAIN CB16).
SQ   SEQUENCE   1276 AA;  146871 MW;  C1EC50F46C8233E2 CRC64;

  Query Match             24.0%;  Score 612;  DB 1;  Length 1276;
  Best Local Similarity   31.0%;  Pred. No. 3.2e-29;
  Matches  146;  Conservative   81;  Mismatches  176;  Indels   68;  Gaps   11;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :::|  : | | ::|: : ||||  :   |:|:||| | : || :|  | |    |||::
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              |    |  : :  ::| ::::| |::||::  |   |       ||||||: |: :|||
Db        910 S-GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              : |:   | | | ||| | ||: |:||   : : | |||| :||||: :|   :|||| |
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              |:|:|| :|| :     | | :    |  : :||:|||| :||:  ::: :|  |  |
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDININVYEGQILRN 1083

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINY---QNLYL 323
              :||:|||||::||:||: |    : ||           || :|     : |     ||
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI----------APESNV-LVLVQYPDRSKLYT 1131

Qy        324 GLRFIIKKASN----SRNINNDNIVREGDYIYLNIDNISDESYRVYYVLVNSKEIQTQLFL 379
              |    ||  |:    || :| |||:                   :::| ||::
Db       1132 GNPITIKSVSDKNPYSRILNGDNII------------------LHMLYNSRKYMI---- 1168

Qy        380 APINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNY- 438
                | | | |   |  :  :   | ::          |:| |   |    |    : ::
Db       1169 --IRDTDTIY-ATQGGECSQNCVYALKLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFR 1225

Qy        439 -------------FCISQWYLRRISENINKLRLGCN--WQFIPVDEGWTE 473
                          |   |      |    | :  |:||  | || |
Db       1226 ENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1276


RESULT 10
BXC1_CLOBO
ID   BXC1_CLOBO     STANDARD;      PRT;  1290 AA.
AC   P18640;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)
DE   (Bontoxilysin C1).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90370487; PubMed=2204031;
RA   Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,
RA   Boquet P., Popoff M.R.;
RT   "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
RL   Nucleic Acids Res. 18:4924-4924(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=91024998; PubMed=2222445;
RA   Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
RA   Yokosawa N., Takeshi K., Syuto B., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for botulinum
RT   type C1 toxin in the C-ST phage genome.";
RL   Biochem. Biophys. Res. Commun. 171:1304-1311(1990).
RN   [3]
RP   SEQUENCE OF 2-25.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=88153072; PubMed=2450068;
RA   Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,
RA   Oguma K.;
RT   "Establishment of a monoclonal antibody recognizing an antigenic site
RT   common to Clostridium botulinum type B, C1, D, and E toxins and
RT   tetanus toxin.";
RL   Infect. Immun. 56:898-902(1988).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94038966; PubMed=7901002;
RA   Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;
RT   "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
RT   cleaving HPC-1/syntaxin.";
RL   EMBO J. 12:4821-4828(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H). The light chain has the pharmacological activity,
CC       while the N- and C-terminal of the heavy chain mediate channel
CC       formation and toxin binding, respectively.
```

```
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C
CC       STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC       BACTERIOPHAGE.
CC   -!- SIMILARITY: Belongs to peptidase family M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66433; CAA47060.1; -.
DR   EMBL; X72793; CAA51313.1; -.
DR   EMBL; X53751; CAA37780.1; -.
DR   EMBL; D90210; BAA14235.1; -.
DR   EMBL; X62389; CAA44263.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    448       BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
FT   CHAIN       449   1290       BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    229    229       BY SIMILARITY.
FT   METAL       232    232       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    452       INTERCHAIN (PROBABLE).
FT   CONFLICT     84     84       P -> T (IN REF. 2).
SQ   SEQUENCE   1290 AA;  148734 MW;  71FBE379F97129E8 CRC64;

  Query Match             22.6%;  Score 575.5;  DB 1;  Length 1290;
  Best Local Similarity   31.7%;  Pred. No. 4.8e-27;
  Matches  151;  Conservative   84;  Mismatches  169;  Indels   73;  Gaps   16;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :::|  : | | :||: : ||||  |   |:|:||| | ::   ||  | |    ||| :
Db        853 NSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQLNPIFPFDFKLGS 912

Qy         87 S--ENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSG 144
              |  :   :   |:: :||:||:::|||:||:|  |: :|         ||||  :||:||
Db        913 SGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN------LPGYTIIDSVKNNSG 966

Qy        145 WKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYING 204
              | : |  | :::||       :|| | | | :|   | |||| :|:||: :||  |||||
Db        967 WSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGY-NKWFFVTVTNNMMGNMKIYING 1025

Qy        205 SLKKSEKILNLDRINSSNDIDFKLINCTDTTKF--------VWIKDFNIFGRELNATEVS 256
              |  : |:  |  || |  | |::    ||           :||:|| || :||:  :::
Db       1026 KLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDIN 1085

Qy        257 SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTN-FNNAA 315
               |:     || :||:||| |||: :||: |    | |:   |:  :  |    | ||
Db       1086 ILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNE-- 1143

Qy        316 INYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNI-----------------DNI 357
                 | : |||:    |   ||  || || :| ::                     ||
Db       1144 --------GYKIIIKRI---RGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNH 1192

Qy        358 SDES-YRVYVLVNSKEIQTQLF--LAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTK 414
              | |  | : :   :|:|    :   : |:|:  |:|   ||     |: :| : :
Db       1193 STEDIYAIGLREQTKDINDNIIFQIQPMNN--TYYYASQI----FKSNFNGENIS----- 1241

Qy        415 TFGLFGIGKFVKDYGYVWDTYDNYF--CISQWYLRRISENINKLRLGCNWQFIPVDE 469
                :  || :    |  |  : ||    : |    : |: :    :| |:|| |
Db       1242 --GICSIGTYRFRLGGDWYRH-NYLVPTVKQGNYASLLESTS-----THWGFVPVSE 1290


RESULT 11
BXEN_CLOBU
ID   BXEN_CLOBU     STANDARD;      PRT;  1162 AA.
AC   Q06366;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=93360835; PubMed=8355622;
RA   Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA   Takeshi K., Ohyama T., Isogai E., Isogai H.;
RT   "Similarity in nucleotide sequence of the gene encoding nontoxic
RT   component of botulinum toxin produced by toxigenic Clostridium
RT   butyricum strain BL6340 and Clostridium botulinum type E strain
RT   Mashike.";
RL   Microbiol. Immunol. 37:395-398(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12739; BAA02231.1; -.
DR   PIR; I40817; I40817.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136829 MW;  C86E9BE932DA78E4 CRC64;

  Query Match              8.8%;  Score 224;  DB 1;  Length 1162;
  Best Local Similarity   23.0%;  Pred. No. 4.2e-06;
  Matches   93;  Conservative   77;  Mismatches  164;  Indels   70;  Gaps   17;

Qy         58 DSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVR 116
              |:||     :    |: :   | |     |  : | ||       |   : : :|||| ||:|
Db        817 DTSGKNTLVEYPKDIGLVYGINNNAIHLTGA-NQNIKFTNDYF--ENGLTNNFSIYFWLR 873

Qy        117 TPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKD 176
                | |::  |       |   ::: ||::  : | ::: :|| |   |:|:
Db        874 --NLNQNTIKSKL------IGSKEDNCGWEIYFENNGLVFNIIDSNGNEKNIYL-----S 920

Qy        177 NISDYINKWFSITITNDRL-GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTT 235
              |||:    |  | |: :||     |:|:  |  :| |  :  | ||:      :|:
Db        921 NISN--KSWHYIVISINRLKDQLLIFIDNILVANEDIKEILNIYSSD-----IISLLSDN 973

Qy        236 KFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK 295
              |:|:  ::  : :|: |: : |:   : : :::|    |:|:  | |||     | |
Db        974 NNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEIAI- 1032

Qy        296 YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID 355
               :|      :| |  ::|:: |    |          : | |   : :  : |:  :|
Db       1033 --NKIEQNNNIYLSNNNENSLNFKPLKFKLL--------NTNPNKQYVQKWDEVIFSVLD 1082
```

```
Qy        356 ------NISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQILC 409
                    :|| :: |: :: |    :| :    ||:|  |                ||
Db       1083 GTEKYLDISIDNNRIQLVDNKNNAKTFI----INND-IFIS------------NCL--- 1121

Qy        410 EKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ----WYLRRI 449
                | |:    :   :|:  | |    |     | :    | |: |
Db       1122 ---TLTYNNVNVYLSIKNQDYNWVICDLNHDIPKKSYLWILKNI 1162


RESULT 12
BXEN_CLOBO
ID   BXEN_CLOBO     STANDARD;      PRT;  1162 AA.
AC   P46082;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type E, nontoxic component.
GN   ENT-120.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mashike;
RX   MEDLINE=93195515; PubMed=8450310;
RA   Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT   "The complete nucleotide sequence of the gene encoding the nontoxic
RT   component of Clostridium botulinum type E progenitor toxin.";
RL   J. Gen. Microbiol. 139:79-86(1993).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12697; BAA02194.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1162 AA;  136856 MW;  96468EDDDAE0F39D CRC64;

  Query Match             8.3%;  Score 212;  DB 1;  Length 1162;
  Best Local Similarity   23.0%;  Pred. No. 2.2e-05;
  Matches   96;  Conservative   75;  Mismatches  149;  Indels   98;  Gaps   21;

Qy         58 DSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVR 116
              |:||    :    |: :   | |     | : | ||       |   : : :|||| ||:|
Db        817 DTSGKNTLVEYPKDIGLVYGINNNAIHLTGA-NQNIKFTNDYF--ENGLTNNFSIYFWLR 873

Qy        117 TPKYNNNDIQTYLQNEYTIISCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKD 176
                |   | |:: |       |   ::: ||::  : : :::.:|| |    |:|:
Db        874 NLK--QNTIKSKL------IGSKEDNCGWEIYFENDGLVFNIIDSNGNEKNIYL-----S 920

Qy        177 NISDYINKWFSITITNDRL-GNANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTT 235
              |||:    |  | |: :||     |:|:  |  :| |  :  | ||:     :|:
Db        921 NISN--KSWHYIVISINRLKDQLLIFIDNILVANEDIKEILNIYSSD-----IISLLSDN 973

Qy        236 KFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIK 295
                |:|:  ::  : :|: |: : |:   : : :::|    |:|:  | |||      | |
Db        974 NNVYIEGLSVLNKTINSNEILTDYFSDLNNSYIRNFDEEILQYNRTYELFNYVFPEIAIN 1033

Qy        296 YFSKASMGETAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNIN------------NDN 342
                :             ||      |:|| :       :|  |:|             |
Db       1034 KIEQ------------NN------NIYLSI-------NNENNLNFKPLKFKLLNTNPNKQ 1068

Qy        343 IVREGDYIYLNI-------DNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIK 395
                |:: | :  ::         :||  : |:  ||::|    |:|:  ||:|  |
Db       1069 YVQKWDEVIFSVLDGTEKYLDISTTNNRIQ-LVDNKN-NAQIFI--INND-IFIS----- 1118

Qy        396 KYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ----WYLRRI 449
                      ||       | |:    :   :|:  | |    |     | :    | |: |
Db       1119 --------NCL------TLTYNNVNVYLSIKNQDYNWVICDLNHDIPKKSYLWILKNI 1162


RESULT 13
BXCN_CLOBO
ID   BXCN_CLOBO     STANDARD;      PRT;  1196 AA.
AC   P46081;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Botulinum neurotoxin type C1, nontoxic component.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Type C Stockholm / C-ST;
RX   MEDLINE=92231894; PubMed=1567404;
RA   Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT   "The complete nucleotide sequence of the gene coding for the
RT   nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT   progenitor toxin.";
RL   Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC   -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC       TOXICITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62389; CAA44262.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000395; Peptidase_M27.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1196 AA;  138740 MW;  4BD5956274D7F9C3 CRC64;

  Query Match             7.7%;  Score 197;  DB 1;  Length 1196;
  Best Local Similarity   24.2%;  Pred. No. 0.00018;
  Matches   90;  Conservative   65;  Mismatches  121;  Indels   96;  Gaps   21;

Qy         79 NGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISC 138
              ||:     | |  |    :||   : : :||:| ||:|      || :| |      |
Db        873 NGESLYLKSPNETI-KFSNKFFT-NGLTNNFTICFWLRFT--GKNDDKTRL------IGN 922

Qy        139 IKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRL-GN 197
                |: ||::  : | ::: :|| |   :|::    | ||       |: |:|: |||
Db        923 KVNNCGWEIYFEDNGLVFEIIDSNGNQESVYLSNIINDN-------WYYISISVDRLKDQ 975

Qy        198 ANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSS 257
                |:||     :  | :  | |:|     :|:  :    ::::: ::    : : ||
Db        976 LLIFINDKNVANVSIDQILSIYSTN-----IISLVNKNNSIYVEELSVLDNPITSEEVIR 1030

Qy        258 LYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAIN 317
                |:     : ::|   : | |:  | |:|      |:  | : |: |     | || :
Db       1031 NYFSYLDNSYIRDSSKSLLEYNKNYQLYN------YV--FPETSLYE----VNDNNKS-- 1076

Qy        318 YQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNID---------------------- 355
```

```
                 || |:       |:  ||  ::     :  :|||
Db       1077 ----YLSLK-------NTDGINISSV----KFKLINIDESKVYVQKWDECIICVLDGTEK 1121

Qy        356 --NISDESYRVYVLVNSKE------IQTQLFLAPINDDPTF-YDVLQIKKYYEKT----T 402
              :|| |: |:  ||:||:      : | ||     |  || |:      ||:  :
Db       1122 YLDISPENNRIQ-LVSSKDNAKKITVNTDLFRP---DCITFSYN----DKYFSLSLRDGD 1173

Qy        403 YNCQILCEKDTK 414
              ||  | |  : |
Db       1174 YNWMI-CNDNNK 1184


RESULT 14
TOXA_CLODI
ID   TOXA_CLODI     STANDARD;      PRT;  2710 AA.
AC   P16154;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Toxin A.
GN   TOXA OR TCDA.
OS   Clostridium difficile.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1496;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90221894; PubMed=2109310;
RA   Sauerborn M., von Eichel-Streiber C.;
RT   "Nucleotide sequence of Clostridium difficile toxin A.";
RL   Nucleic Acids Res. 18:1629-1630(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RX   MEDLINE=90129305; PubMed=2105276;
RA   Dove C.H., Wang S.Z., Price S.B., Phelphs C.J., Lyerly D.M.,
RA   Wilkins T.W., Johnson J.L.;
RT   "Molecular characterization of the Clostridium difficile toxin A
RT   gene.";
RL   Infect. Immun. 58:480-488(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI 10463;
RA   von Eichel-Streiber C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Only after the enteral delivery of the enterotoxin A may
CC       the characteristic disease called pseudomembranous colitis be
CC       induced.
CC   -!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC       REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC       DIFFERENT OLIGOPEPTIDES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51797; CAA36094.1; -.
DR   EMBL; M30307; AAA23283.1; -.
DR   EMBL; X92982; CAA63564.1; -.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR007577; Gly_transf_sug.
DR   Pfam; PF01473; CW_binding_1; 28.
DR   Pfam; PF04488; Gly_transf_sug; 1.
KW   Toxin; Enterotoxin.
SQ   SEQUENCE   2710 AA;  308052 MW;  0A6E52CE84C14421 CRC64;

  Query Match             6.2%;  Score 158;  DB 1;  Length 2710;
  Best Local Similarity   20.3%;  Pred. No. 0.098;
  Matches  107;  Conservative   94;  Mismatches  164;  Indels  162;  Gaps   28;

Qy         35 NNYISNISSNAI---LSLSYRG-----GRLIDSSGYGATMNVGSD--VIFN--------D 76
              |  :  |::| |   || |: |       |: |      :|:  |   |||
Db       1298 NFIMPTITTNEIRNKLSYSFDGAGGTYSLLLSSYPISTNINLSKDDLWIFNIDNEVREIS 1357

Qy         77 IGNGQF---KLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEY 133
              | ||     ||     | |  :::| ::       | :|:|         : ||    :: |
Db       1358 IENGTIKKGKLIKDVLSKIDINKNKLII-----GNQTIDF--------SGDIDN--KDRY 1402

Qy        134 TIISCIKNDSGWKVSIKGNRIIWTLIDVN--AKSKSIFF----EYSIKDNISDYINKWFS 187
               ::|  :|   |:|:       :|::| ||| |:       | |  |:|: | |  :
Db       1403 IFLTCELDD---KISL--------IIEINLVAKSYSLLLSGDKNYLI-SNLSNTIEKINT 1450

Qy        188 ITITNDRLG------NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIK 241
              : : :  :       : | |     | |:| :   : :| | ::|      |:|     |
Db       1451 LGLDSKNIAYNYTDESNNKYFGAISKTSQKSIIHYKKDSKNILEF----YNDSTLEFNSK 1506

Qy        242 DF-----NIFGRELNATEVSSLYWIQSSTNTLKDF-----WGNPLRYDTQYYLFNQGMQN 291
              ||     |:| :: :   ::  |:: ::|:   ||       | :: :  |   |: : :
Db       1507 DFIAEDINVFMKD-DINTITGKYYVDNNTDKSIDFSISLVSKNQVKVNGLY--LNESVYS 1563

Qy        292 IYIKYFSKASMGETAPRTNFNNAAINYQNLYL-----------GLRFIIKK-------- 331
              |:  | | | |        ::   |: ||:|            : |:| |
Db       1564 SYLD-FVKNSDG-------HHNTSNFMNLFLDNISFWKLFGFENINFVIDKYFTLVGKT 1614

Qy        332 --------ASNSRNI-----------------NNDNIVRE-------GDYIYLNIDNI 357
                     |::||                  |  |:| |       |: |  ::|
Db       1615 NLGYVEFICDNNKNIDIYFGEWKTSSSKSTIFSGNGRNVVVEPIYNPDTGEDISTSLDFS 1674

Qy        358 SDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEK------TTYNCQILCEK 411
               :  | :   :|   |   |: : || :  :|      :|| :       |:: ::
Db       1675 YEPLYGIDRYINKVLIAPDLYTSLININTNYYS----NEYYPEIIVLNPNTFHKKVNINL 1730

Qy        412 DTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRL 458
              |: :|            | | |  : | : ::     : : |:|:
Db       1731 DSSSF-----------EYKWSTEGSDFILVRYLEESNKKILQKIRI 1765


RESULT 15
YCF2_MARPO
ID   YCF2_MARPO     STANDARD;      PRT;  2136 AA.
AC   P09975;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN   YCF2.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC   Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
CC   -!- SIMILARITY: Belongs to the ycf2 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04465; CAA28078.1; -.
```

```
DR   PIR; S01591; A05037.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
KW   Chloroplast; Hypothetical protein.
SQ   SEQUENCE   2136 AA;  259911 MW;  5BD170C9CCF61197 CRC64;

  Query Match             6.0%;  Score 152;  DB 1;  Length 2136;
  Best Local Similarity   24.5%;  Pred. No. 0.17;
  Matches  103;  Conservative   70;  Mismatches  132;  Indels  116;  Gaps   29;

Qy         79 NGQFKLNNS--ENSNITAHQSKFVVYDSMFDNFSINFWVRTPK-YNNNDIQTYLQNEYTI 135
              ||::|: :   :|      : |          ||  |  ::    ||::    ||||:|
Db        652 NGEYKIESQILQNEKELNKKRK--------KNFQFNPNIKILSFYNSSKKNIYLQNKYFF 703

Qy        136 ISCIKNDS--GW-KVSIK-------GNRIIWTLIDVNAKSKSIFFEYSIKDNISD--YIN 183
                 : |:    | |:| |         |:|||       |    || :| |::: |  : |
Db        704 NKNLINNKLITWKKISNKLVISNSEYNKIIWN------KKNMKFFSFS-KNSVLDTFFFN 756

Qy        184 K--WFSITITNDRLGNANIYINGSLKKSEKILNLDRI--NSSNDIDFKL------INCTD 233
              |  :  ||:  |:|    :    : :: :||||   :  || |    |:      ||
Db        757 KKSFNIITVIFDKLKKIQL----NFQEIQKILNCFSLFFNSKNIKKTKIFKNSYFINENL 812

Qy        234 TTKFVW-IKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWGNPLR---YDTQYYLFN--- 286
              || | :  |:|||  ||  :|:::       : |  |:   |    |  :  |||
Db        813 TTTFSFNDKEFNIFFLELFISEINN--------DFLMRFFKKYLYYRIYKDKEILFNPIE 864

Qy        287 --QGMQNIYIKYFSKASMGE-----TAPRTNFNNAAINYQNLYLGLRFIIKKASNSRNIN 339
                | :||    :| |  :          |  |:||            |||       : |
Db        865 NRQLLQN----FFEKTKILTFIDFLQDPELNYNN-----------RFIFH--LEKKTIK 906

Qy        340 NDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND---DPTFYDVLQIKK 396
              |:|:      :|| :  |  :  | ::]:|  ||::  |:   |:        :|| :|
Db        907 NNNL------LYLRLLKIFLKDKRNFLLIN--EIKS--FIEKKNNLFIKSQLSNVLLVKN 956

Qy        397 YYE--KTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYF---CISQWYLRRISE 451
              |:     :|   | :|:            |:   : :   |||    : : ||: ::
Db        957 SYKFFDNIFNFHFLKQKE-----------KNIEIILNN-QNYFEKSLLKKTYLKNLNL 1002

Qy        452 N 452
              |
Db       1003 N 1003


Search completed: September 10, 2004, 03:36:05
Job time : 8.77122 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2004, 03:33:56 ; Search time 21.6421 Seconds
(without alignments)
6895.826 Million cell updates/sec

Title: US-10-729-039-79
Perfect score: 2550
Sequence: 1 MGHHHHHHHHHHSSGHIEGR.....NKLRLGCNWQFIPVDEGWTE 473

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1202.5	47.2	1291	2	Q933K0	Q933k0 clostridium
2	1201.5	47.1	1291	2	Q9ZAJ8	Q9zaj8 clostridium
3	1198.5	47.0	1291	2	Q93G71	Q93g71 clostridium
4	1196.5	46.9	1291	2	Q8GR96	Q8gr96 clostridium
5	1177.5	46.2	1291	2	Q08077	Q08077 clostridium
6	1152.5	45.2	441	2	Q9X708	Q9x708 clostridium
7	744.5	29.2	1268	2	Q45851	Q45851 clostridium
8	742.5	29.1	1251	2	Q9K395	Q9k395 clostridium
9	736	28.9	1252	2	Q8KZM3	Q8kzm3 clostridium
10	736	28.9	1255	2	Q9FAR6	Q9far6 clostridium
11	724.5	28.4	1278	2	Q57236	Q57236 clostridium
12	668	26.2	1280	2	Q9ZAJ5	Q9zaj5 clostridium
13	632.5	24.8	451	2	Q9LA13	Q9la13 clostridium
14	607.5	23.8	1275	12	Q9QTG7	Q9qtg7 clostridium
15	607	23.8	1280	2	Q841S3	Q841s3 clostridium
16	606	23.8	1280	2	Q45849	Q45849 clostridium
17	606	23.8	1280	2	Q9LBS7	Q9lbs7 clostridium
18	605.5	23.7	1310	2	Q93N27	Q93n27 clostridium
19	575.5	22.6	1291	2	Q93HT3	Q93ht3 clostridium
20	543.5	21.3	1285	2	Q45967	Q45967 clostridium
21	542.5	21.3	1285	2	Q9LBR1	Q9lbr1 clostridium
22	461	18.1	361	2	Q45848	Q45848 clostridium
23	461	18.1	361	2	Q45846	Q45846 clostridium
24	321	12.6	367	2	Q45861	Q45861 clostridium
25	321	12.6	367	2	Q45862	Q45862 clostridium
26	229.5	9.0	1198	2	Q45893	Q45893 clostridium
27	227.5	8.9	1198	2	O69277	O69277 clostridium
28	226.5	8.9	1198	2	O06018	O06018 clostridium
29	219.5	8.6	1160	2	O87710	O87710 clostridium
30	218.5	8.6	1197	2	P71117	P71117 clostridium
31	218.5	8.6	1197	2	O33871	O33871 clostridium
32	213.5	8.4	1197	2	Q45888	Q45888 clostridium
33	213	8.4	1197	2	Q9ZAJ9	Q9zaj9 clostridium
34	203.5	8.0	1162	2	Q9ZAJ6	Q9zaj6 clostridium
35	201.5	7.9	1161	2	O69276	O69276 clostridium
36	200.5	7.9	1161	2	Q45891	Q45891 clostridium
37	198.5	7.8	1193	2	Q45880	Q45880 clostridium
38	197	7.7	1196	2	Q45916	Q45916 clostridium
39	197	7.7	1196	2	Q53550	Q53550 clostridium
40	197	7.7	1196	2	Q93HT4	Q93ht4 clostridium
41	197	7.7	1196	9	Q9ZX77	Q9zx77 clostridium
42	197	7.7	1196	9	Q38197	Q38197 clostridium
43	193.5	7.6	3063	5	Q8I284	Q8i284 plasmodium
44	193	7.6	1196	2	Q9LBR2	Q9lbr2 clostridium
45	192.5	7.5	1193	2	P71107	P71107 clostridium

ALIGNMENTS

RESULT 1
Q933K0
ID Q933K0 PRELIMINARY; PRT; 1291 AA.
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=593, and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;

Query Match 47.2%; Score 1202.5; DB 2; Length 1291;

```
  Best Local Similarity   50.9%;  Pred. No. 3.5e-57;
  Matches  230;  Conservative   94;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :||||::|| | |:| :| ||:| ||  :||| ||||| : |    | ||     |||| :
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 903

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: :::: ||| :||::||:| ||| |: || |: |||||:|:||:|||| 
Db        904 SANSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|| |||||||:| | ||:|||||||::||:|||:|| :||||:   || ||||| | 
Db        964 ISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS-DNAKIYINGKL 1022

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              :    | ::  : ::::| |||     | |:|:|:| |:||  ||: : :  :| ||| : 
Db       1023 ESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSE 1082

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||||  |: :||:|| | :| |||    :|:||   |: :|  :  |||::||:| 
Db       1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIG 1142

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||:    |: : :|||||:| 
Db       1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYMYKYFKKEEEKLFLAPISD 1200

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
              ||: :|||:| |: ||:||:| :||   |  || || :| :  | |:  | :|||| 
Db       1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFKEYKDYFCI 1259

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :       :||||||||| ||||| 
Db       1260 SKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 2
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BonT protein.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281;
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;

  Query Match           47.1%;  Score 1201.5;  DB 2;  Length 1291;
  Best Local Similarity   50.9%;  Pred. No. 4e-57;
  Matches  230;  Conservative   94;  Mismatches  117;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :||||::|| | |:| :| ||:| ||  :||| ||||| : |    | ||     |||| :
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 903

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: :::: ||| :||::||:| ||| |: || |: |||||:|:||:|||| 
Db        904 SANSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|| |||||||:| | ||:|||||||::||:|||:|| :||||:   || ||||| | 
Db        964 ISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS-DNAKIYINGKL 1022

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              :    | ::  : ::::| |||     | |:|:|:| |:||  ||: : :  :| ||| : 
Db       1023 ESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSE 1082

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||||  |: :||:|| | :| |||    :|:||   |: :|  :  |||::||:| 
Db       1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSKYINYRDLYIG 1142

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||:    |: : :|||||:| 
Db       1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYMYKYFKKEEEKLFLAPISD 1200

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
              ||: :|||:| |: ||:||:| :||   |  || || :| :  | |:  | :|||| 
Db       1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFKEYKDYFCI 1259

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :       :||||||||| ||||| 
Db       1260 SKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 3
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
```

```
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;

  Query Match             47.0%;  Score 1198.5;  DB 2;  Length 1291;
  Best Local Similarity   50.7%;  Pred. No. 5.8e-57;
  Matches  229;  Conservative   94;  Mismatches  118;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :||||::|| | |:| :| ||:| ||  :||| ||||| : |    |  ||    |||| :
Db        846 NTILIEIFNKYNSDILNNIILNLRYRDNKLIDLSGYGAKVEVYDGVKLND--KNQFKLTS 903

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: ::::||| :||::||:| ||| |: || |: |||||:|:||:||||
Db        904 SANSKIRVIQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|| |||||||:| | ||:|||||||::||:|||:|| :||||:   || ||||| |
Db        964 ISIRGNMIIWTLIDINGKIKSVFFEYSIKEDISEYINRWFFVTITNNS-DNAKIYINGKL 1022

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              :    | ::  : ::::| |||    | |:|:|:| |:||  ||: : :  :| ||| :
Db       1023 ESHIDIRDIREVIANDEIIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEIYKIQSYSE 1082

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||| |: :||:|| | :| |||    : :||   |: :|  :  |||::||:|
Db       1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLYIG 1142

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  |:: | :|||:    |: : :||||||:|
Db       1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFNLNQE-WRVYIYKYFKKEEEKLFLAPISD 1200

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
                ||: :|||:| |: ||:||:| :||    |   || || :| :  | |:  | :||||
Db       1201 SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVFKEYKDYFCI 1259

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :       :|||||||| ||||||
Db       1260 SKWYLKEVKRKPYNSKLGCNWQFIPKDEGWTE 1291


RESULT 4
Q8GR96
ID   Q8GR96      PRELIMINARY;      PRT;  1291 AA.
AC   Q8GR96;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
GN   BONTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA   Nakamura S., Mukamoto M., Kozaki S.;
RT   "Clostridium botulinum type B neurotoxin associated with infant
RT   botulism.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB084152; BAC22064.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150574 MW;  0227CAEF4F58504D CRC64;

  Query Match             46.9%;  Score 1196.5;  DB 2;  Length 1291;
  Best Local Similarity   51.3%;  Pred. No. 7.4e-57;
  Matches  232;  Conservative   84;  Mismatches  125;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :||||::|| | | :| ||:| ||   ||| ||||| : |    |  ||    |||| :
Db        846 NTILIEIFNKYNSEILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELND--KNQFKLTS 903

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              | || |   |:: ::::||| :||::||:| ||| |: || |: |||||:||||:||||
Db        904 STNSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              :||:|||||||| |:| |:||:|||||:::|||||:|| :||||:   || ||||| |
Db        964 ISIRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNS-DNAKIYINGKL 1022

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              : :  | ::  : :: :| |||    | |:|:|:| |:||  ||: : :  :| ||| :
Db       1023 ESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSE 1082

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
              |||||||| |: :||:|| | :| |||    :|:||   |: :|  :  |||:|||:|
Db       1083 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIG 1142

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :|||:: |||::| ||:|||: |||||:  | |:  :|||   : || : :|||| | |
Db       1143 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFN-SNREWRVYAYKDFKEEEKKLFLANIYD 1200

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
                ||  :|||:| |: ||:||:| :||    |   || || :| :  | |   | |||||
Db       1201 SNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVLKDYKNYFCI 1259

Qy        442 SQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |:|||: :         |||||||| |||| |
Db       1260 SKWYLKEVKRKPYNPNLGCNWQFIPKDEGWIE 1291


RESULT 5
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BoNT/B.
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Eklund 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
DR   EMBL; X71343; CAA50482.1; -.
DR   PIR; I40631; I40631.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;

  Query Match             46.2%;  Score 1177.5;  DB 2;  Length 1291;
  Best Local Similarity   50.0%;  Pred. No. 7.9e-56;
  Matches  225;  Conservative   94;  Mismatches  120;  Indels   11;  Gaps    7;

Qy         29 ILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNNSE 88
Db        848 ILIKIFNKYNSEILNNIILNLRYRDNNLIDLSGYGAKVEVYDGVKLND--KNQFKLTSSA 905

Qy         89 NSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWKVS 148
Db        906 DSKIRVTQNQNIIFNSMFLDFSVSFWIRIPKYRNDDIQNYIHNEYTIINCMKNNSGWKIS 965

Qy        149 IKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSLKK 208
Db        966 IRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNN-LDNAKIYINGTLES 1024

Qy        209 SEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTL 268
Db       1025 NMDIKDIGEVIVNGEITFKLDGDVDRTQFIWMKYFSIFNTQLNQSNIKEIYKIQSYSEYL 1084

Qy        269 KDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLGLR 326
Db       1085 KDFWGNPLMYNKEYYMFNAGNKNSYIKLVKDSSVGEILIRSKYNQNSNYINYRNLYIGEK 1144

Qy        327 FIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDDP 386
Db       1145 FIIRRESNSQSI-NDDIVRKEDYIHLDL-VLHHEEWRVYAYKYFKEQEEKLFLSIISDSN 1202

Qy        387 TFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCISQ 443
Db       1203 EFYKTIEIKEYDEQPSYSCQLLFKKDEESTDDIGLIGIHRFYES-GVLRKKYKDYFCISK 1261

Qy        444 WYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db       1262 WYLKEVKRKPYKSNLGCNWQFIPKDEGWTE 1291


RESULT 6
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type B (Fragment).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;

  Query Match             45.2%;  Score 1152.5;  DB 2;  Length 441;
  Best Local Similarity   50.7%;  Pred. No. 5.6e-55;
  Matches  225;  Conservative   84;  Mismatches  124;  Indels   11;  Gaps    7;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
Db          4 NTILIEIFNKYNSKILNNIILNLRYRDNNLIDLSGYGANVEVYDGVELND--KNQFKLTS 61

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
Db         62 STNSEIRVTQNQNIIFNSMFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCIKNNSGWK 121

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
Db        122 ISIRGNRIIWTLTDINGKTKSVFFEYSIREDISDYINRWFFVTITNNS-DNAKIYINGKL 180

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
Db        181 ESNIDIKDIGEVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIKEIYKIQSYSE 240

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAA--INYQNLYLG 324
Db        241 YLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSSVGEILTRSKYNQNSNYINYRNLYIG 300

Qy        325 LRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
Db        301 EKFIIRRKSNSQSI-NDDIVRKEDYIYLDFFN-SNREWRVYAYKDFKEEEKKLVLANIYD 358

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKD---TKTFGLFGIGKFVKDYGYVWDTYDNYFCI 441
Db        359 SNEFYKTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYES-GIVLKDYKNYFCI 417

Qy        442 SQWYLRRISENINKLRLGCNWQFI 465
Db        418 SKWYLKEVKRKPYNPNLGCNWQFI 441


RESULT 7
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin type F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   PIR; S33411; S33411.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
```

```
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;

  Query Match           29.2%;  Score 744.5;  DB 2;  Length 1268;
  Best Local Similarity   34.1%;  Pred. No. 2.2e-32;
  Matches  162;  Conservative   85;  Mismatches  151;  Indels   77;  Gaps   12;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | ||| :   :   |  ::||:: |   | |||||||: :::  |:          || : :
Db        843 DKILIHILIRFYKRIIDSSILNMKYENNRFIDSSGYGSNISINGDIYIYSTNRNQFGIYS 902

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
              |  | :   |:  ::|:| : |||::|||| |||||       | ||||||:|:: |:|||
Db        903 SRLSEVNITQNNTIIYNSRYQNFSVSFWVRIPKYNN---LKNLNNEYTIINCMRNNNSGW 959

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |:|:  | ||||| |    :: : | |:   :|||||||  :||||:|||:: :||||:
Db        960 KISLNYNNIIWTLQDTTGNNQKLVFNYTQMIDISDYINKWTFVTITNNRLGHSKLYINGN 1019

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : ||||  |:  ::| ||:: | | |::| |: | ||  ||: ||: :||  :  :
Db       1020 LTDQKSILNLGNIHVDDNILFKIVGCND-TRYVGIRYFKIFNMELDKTEIETLYHSEPDS 1078

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIK-------------YFSKASMGETAPRTNFN 312
                |||||| | |: :||| |    |: :                :|| ::        |:
Db       1079 TILKDFWGNYLLYNKKYYLLNLLKPNMSVTKNSDILNINRQRGIYSKTNI--------FS 1130

Qy        313 NAAINYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKE 372
              ||      || |:  ||:|  ::   | || ||: | :|:|: : : | |::|  |::
Db       1131 NA-----RLYTGVEVIIRKVGSTDTSNTDNFVRKNDTVYINVVDGNSE-YQLYADVSTSA 1184

Qy        373 IQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQIL---------CEKDTKT-----FGL 418
              ::  : |  |::               : ||   :          |  : ||      ||
Db       1185 VEKTIKLRRISN---------------SNYNSNQMIIMDSIGDNCTMNFKTNNGNDIGL 1228

Qy        419 FGIGKFVKDYGYVWDTYDNYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |              : |    | || : |  |      || | ||  : || |
Db      .1229 LGF------------HLNNLVASSWYYKNIRNNTR--NNGCFWSFISKEHGWQE 1268


RESULT 8
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
```

```
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;

  Query Match             29.1%;  Score 742.5;  DB 2;  Length 1251;
  Best Local Similarity   35.9%;  Pred. No. 2.8e-32;
  Matches  161;  Conservative   86;  Mismatches  169;  Indels   33;  Gaps   12;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  || :   | |:::|:: |:  : :|:||| : :|:  ::          || : |
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGEIFIYPTNKNQFTIFN 892

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
              |: | :   |: :::||: : ||||:|||| | |:|  :   : ||||||:|:: |:|||
Db        893 SKPSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--INNEYTIINCMRDNNSGW 950

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |||:  | ||||| |    :: : |:|   : |||||||  :|||||||:: :||||
Db        951 KVSLNHNEIIWTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRLGDSKLYINGH 1010

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : ||||  |: |::| ||::||:  |::: |: |||| :||: ||: :||  : :|
Db       1011 LIDQKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYSNEPNT 1069

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325
              | ||||||| | ||  ||| |    | :|       |   | :  :    || |:
Db       1070 NILKDFWGNYLLYDKGYYLLNVLKPNNFI----DRRKDSTLSINNIRSTILLANRLYSGI 1125

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLN-IDNISDESYRVYVLVNSKEIQTQLFLAPIND 384
              :  |:: ::|    ||  ||: | :|:| | | |  || :|   |: :   :  :   :
Db       1126 KVKIQRVNDSS--TNDRFVRKNDQVYINYISNSS--SYSLYADTNTTD--KEKTIKSSSS 1179

Qy        385 DPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQW 444
              |   |: :       | | :     :   ||          |: ||      | |
Db       1180 GNRFNQVVVMNSVGNNCTMNFK---NNNGNNIGLL---------GFKADT----VVASTW 1223

Qy        445 YLRRISENINKLRLGCNWQFIPVDEGWTE 473
              |   : :: |    || | ||  : || |
Db       1224 YYTHMRDHTNS--NGCFWNFISEEHGWQE 1250


RESULT 9
Q8KZM3
ID   Q8KZM3      PRELIMINARY;      PRT;  1252 AA.
AC   Q8KZM3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 5262;
RA   Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA   Nakamura S., Karasawa T., Kozaki S.;
RT   "Clostridium butyricum bont/E gene for type E botulinum toxin,
RT   complete cds.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB088207; BAC05434.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1252 AA;  143510 MW;  41B633BB744D3B41 CRC64;

  Query Match             28.9%;  Score 736;  DB 2;  Length 1252;
  Best Local Similarity   34.8%;  Pred. No. 6.4e-32;
  Matches  156;  Conservative   91;  Mismatches  171;  Indels   30;  Gaps   10;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  || :   | |:::|:: |:  : :|:||| : :|:  ||          || : |
Db        833 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 892

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
              : | :   |: :::||: : ||||:|||| | |:|  :   : ||||||:|:: |:|||
Db        893 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 950

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |||:  | ||||| | :  :: : | |    : |||||||  :|||||||:: :||||:
Db        951 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1010

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : ||||  |: |::| ||::||:  |::: |: |||| :||: ||: :||  : :
Db       1011 LIDKKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1069

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325
              | ||||||| | || :||| |    | :|       |   | :  :    || |:
Db       1070 NILKDFWGNYLLYDKEYYLLNVLKPNNFI----NRRTDSTLSINNIRSTILLANRLYSGI 1125

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 385
              :  |:: :||    |||:||: | :|:|   :: ::: : :  ::      : : :
Db       1126 KVKIQRVNNSS--TNDNLVRKNDQVYINF--VASKTHLLPLYADTATTNKEKTIKISSSG 1181

Qy        386 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 445
              |   |: :       | | :     :   ||          |: ||      | ||
Db       1182 NRFNQVVVMNSVGNNCTMNFK---NNNGNNIGLL---------GFKADT----VVASTWY 1225

Qy        446 LRRISENINKLRLGCNWQFIPVDEGWTE 473
              : :| |    |  | ||  : || |
Db       1226 YTHMRDNTNS--NGFFWNFISEEHGWQE 1251


RESULT 10
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type E botulinum toxin.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;

  Query Match           28.9%;  Score 736;  DB 2;  Length 1255;
  Best Local Similarity   34.8%;  Pred. No. 6.4e-32;
  Matches  156;  Conservative   91;  Mismatches  171;  Indels   30;  Gaps   10;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  || :   | |:::|:: |:  : :|:||| : :|:  ||         || : |
Db        836 DKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYN 895

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
               : | :   |: :::||: : ||||:|||| | |:|  :   : |||||:|:: |:|||
Db        896 DKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVN--VNNEYTIINCMRDNNSGW 953

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |||:  | ||||| | :  :: : | |   : |||||||  :|||||||:: :||||:
Db        954 KVSLNHNEIIWTLQDNSGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGN 1013

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : ||||  |: |::| ||::||:  |::: |: |||| :||: ||: :||  : :
Db       1014 LIDKKSILNLGNIHVSDNILFKIVNCS-YTRYIGIRYFNIFDKELDETEIQTLYNNEPNA 1072

Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQNLYLGL 325
              | ||||||| | || :||| |    | :|          |    |  :  :    || |:
Db       1073 NILKDFWGNYLLYDKEYYLLNVLKPNNFI----NRRTDSTLSINNIRSTILLANRLYSGI 1128

Qy        326 RFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFLAPINDD 385
              :  |:: :||    |||:||: | :|:|   :: ::: : :  ::      :  :   :
Db       1129 KVKIQRVNNSS--TNDNLVRKNDQVYINF--VASKTHLLPLYADTATTNKEKTIKISSSG 1184

Qy        386 PTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYDNYFCISQWY 445
                |  |: :        | | :     :    ||          |:  ||       | ||
Db       1185 NRFNQVVVMNSVGNNCTMNFK---NNNGNNIGLL---------GFKADT----VVASTWY 1228

Qy        446 LRRISENINKLRLGCNWQFIPVDEGWTE 473
                : :| |    |  | ||  : || |
Db       1229 YTHMRDNTNS--NGFFWNFISEEHGWQE 1254


RESULT 11
Q57236
ID   Q57236      PRELIMINARY;      PRT;  1278 AA.
AC   Q57236; Q45863;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Botulinum neurotoxin type F (BONT/F protein).
GN   BONT/F.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10281;
RA   Hutson R.A., Collins M.D.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 635-1000 FROM N.A.
RC   STRAIN=NCTC 1028;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=LANGELAND;
RX   MEDLINE=98404102; PubMed=9732534;
RA   East A.K., Bhandari M., Hielm S., Collins M.D.;
RT   "Analysis of the botulinum neurotoxin type F gene clusters in
RT   proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT   barati.";
RL   Curr. Microbiol. 37:262-268(1998).
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   PIR; S48110; S48110.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;

  Query Match           28.4%;  Score 724.5;  DB 2;  Length 1278;
  Best Local Similarity   35.4%;  Pred. No. 2.8e-31;
  Matches  162;  Conservative   76;  Mismatches  178;  Indels   41;  Gaps   10;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              | |||  ||      |  |:|| : |   : || ||||: :::  ||         || : :
Db        852 DKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYS 911

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIK-NDSGW 145
              |: | :   |:  ::|:  : ||||:|||| ||| |    :  | |||||| ||: |:|||
Db        912 SKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFN---KVNLNNEYTIIDCIRNNNSGW 968

Qy        146 KVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGS 205
              |:|:  |:||||| |    :: : | |:   :|||||||  :||||:||||: |||||:
Db        969 KISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGN 1028

Qy        206 LKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSST 265
              |   : | ||  |: |::| ||:: | | |::| |: | :|  ||  ||: :||  :
Db       1029 LIDEKSISNLGDIHVSDNILFKIVGCND-TRYVGIRYFKVFDTELGKTEIETLYSDEPDP 1087
```

```
Qy        266 NTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINYQ------ 319
Db       1088 SILKDFWGNYLLYNKRYYLLN--------LLRTDKSITQNSNFLNINQQRGVYQKPNIFS 1139

Qy        320 --NLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVN-SKEIQTQ 376
Db       1140 NTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVD-RDVEYRLYADISIAKPEKII 1198

Qy        377 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 436
Db       1199 KLIRTSNSNNSLGQIIVMDSIGNNCTMNFQ---NNNGGNIGLLGF------------HS 1242

Qy        437 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db       1243 NNLVASSWYYNNIRKNTSS--NGCFWSFISKEHGWQE 1277
```

RESULT 12
Q9ZAJ5
ID Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BonT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR002160; Kunitz_legume.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1280 AA; 147487 MW; D0F748976EBC222C CRC64;

Query Match 26.2%; Score 668; DB 2; Length 1280;
Best Local Similarity 34.4%; Pred. No. 3.2e-28;
Matches 157; Conservative 80; Mismatches 182; Indels 38; Gaps 12;

```
Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
Db        852 DKILIIYFNRLYKKIKDSSILDMRYENNKFIDISGYGSNISINGNVYIYSTNRNQFGIYS 911

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKY----NNNDIQTYLQNEYTIISCI-KN 141
Db        912 GRLSEVNIAQNNDIIYNSRYQNFSISFWVTIPKHYRPMNRN-------REYTIINCMGNN 964

Qy        142 DSGWKVSIKGNR---IIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNA 198
Db        965 NSGWKISLRTIRDCEIIWTLQDTSGNKEKLIFRYEELASISDYINKWIFVTITNNRLGNS 1024

Qy        199 NIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSL 258
Db       1025 RIYINGNLIVEKSISNLGDIHVSDNILFKIVGCDDET-YVGIRYFKVFNTELDKTEIETL 1083

Qy        259 YWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSK-ASMGETAPRTNFNNAAIN 317
Db       1084 YSNEPDPSILKDYWGNYLLYNKKYYLFNLLRKDKYITRNSGILNINQQRGVTGGISVFLN 1143

Qy        318 YQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVN-SKEIQTQ 376
Db       1144 YK-LYEGVEVIIRKNAPIDISNTDNFVRKNDLAYINVVDHGVE-YRLYADISITKSEKII 1201

Qy        377 LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD 436
Db       1202 KLIRTSNPNDSLGQIIVMDSIGNNCTMNFQ---NNDGSNIGLLGF------------HS 1245

Qy        437 NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE 473
Db       1246 DDLVASSWYYNHIRRNTSS--NGCFWSFISKEHGWKE 1280
```

RESULT 13
Q9LA13
ID Q9LA13 PRELIMINARY; PRT; 451 AA.
AC Q9LA13;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tetanus toxin (Fragment).
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20886;
RA He H.J., Shi H.J., He Z.Y., Yuan Q.S., Wu X.F.;
RT "Fragment C of Tetanus Toxin.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF154828; AAF73267.1; -.
DR HSSP; P04958; 1A8D.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR002160; Kunitz_legume.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
FT NON_TER 1 1
SQ SEQUENCE 451 AA; 51823 MW; 69A8C5F030E6CD8E CRC64;

Query Match 24.8%; Score 632.5; DB 2; Length 451;
Best Local Similarity 33.9%; Pred. No. 8.6e-27;
Matches 154; Conservative 76; Mismatches 181; Indels 43; Gaps 14;

```
Qy         41 ISSNAILSLSYRGGRLIDSSGYGATMNVGSDV-IFNDIGNGQFKLNNSENSNITAHQSKF 99
Db         17 LKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMD 76

Qy        100 VVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKN-----DSGWKVSIKGNRI 154
Db         77 IEYNDMFNNFTVSFWLRVPKVSASHLEQYDTNEYSIISSMKKYSLSIGSGWSVSLKGNNL 136

Qy        155 IWTLIDVNAKSKSIFFEYSIKDNISDYI-NKWFSITITNDRLGNANIYINGSLKKSEKIL 213
Db        137 IWTLKDSAGEVRQITFR-DLSDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEIT 195

Qy        214 NLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTNTLKDFWG 273
Db        196 GLGAIREDNNITLKLDRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWG 255

Qy        274 NPLRYDTQYYLF--NQGMQNIYIKYFSKASMGETAPRTNFNNAAIN--YQNLYLGLRFII 329
Db        256 NPLRYDTEYYLIPVAYSSKDVQLKNITDYMYLTNAP--SYTNGKLNIYYRRLYSGLKFII 313
```

```
Qy        330 KKASNSRNINNDNIVREGDYIYLNIDNISDESYRVY-----VLVNSKEIQTQLFLAPIND 384
              |: : :  |  |: || ||:| | :   ::|    |         |   |    : ||
Db        314 KRYTPNNEI--DSFVRSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP--- 368

Qy        385 DPTFYDVLQIKKYYEKTTYNCQI-LCEKDTKTFGLFG-----IGKFVKDYGYVWDTYDNY 438
                  |  ::  |  :  ||: |: | :    : || |     ||          |   :
Db        369 GIPLYKKMEAVKLRDLKTYSVQLKLYDDKDASLGLVGTHNGQIGN---------DPNRDI 419

Qy        439 FCISQWYLRRISENINKLRLGCNWQFIPVDEGWT 472
                | ||   : :     | |:| |:| |||||
Db        420 LIASNWYFNHLKDK----TLTCDWYFVPTDEGWT 449


RESULT 14
Q9QTG7
ID   Q9QTG7      PRELIMINARY;      PRT;  1275 AA.
AC   Q9QTG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NTX (Fragment).
GN   NTX.
OS   Clostridium botulinum D bacteriophage.
OC   Viruses.
OX   NCBI_TaxID=29342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1873;
RX   MEDLINE=99017546; PubMed=9802560;
RA   Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA   Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT   "Molecular composition of the 16S toxin produced by a Clostridium
RT   botulinum type D strain, 1873.";
RL   Microbiol. Immunol. 42:599-605(1998).
DR   EMBL; AB012112; BAA75084.1; -.
DR   HSSP; P10845; 3BTA.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
FT   NON_TER    1275   1275
SQ   SEQUENCE   1275 AA;  146742 MW;  3C50F46C8233E2D6 CRC64;

  Query Match           23.8%;  Score 607.5;  DB 12;  Length 1275;
  Best Local Similarity   31.5%;  Pred. No. 6e-25;
  Matches  140;  Conservative   86;  Mismatches  161;  Indels   57;  Gaps    9;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :::|  : | | ::|: : |||| :   |:|:||| | : || :|  | |    |||::
Db        850 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTIYTNDFKLSS 909

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              |   |  : :  ::| ::::| |::||::  |   |       ||||||: |: :||||
Db        910 S-GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIEQNSGWK 963

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              : |:   | | | ||| | ||: |:||   : : | |||| :|||: :|   :|||| |
Db        964 LCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1023

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              |:|:|| :|| :     | | :    | : :||:|||| :||:  ::: :|  |  |
Db       1024 KQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1083

Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINY---QNLYL 323
              :||:|||||::||:||: |    : ||           || :|     : |     ||
Db       1084 VIKDYWGNPLKFDTEYYIINDNYIDRYI----------APESNV-LVLVQYPDRSKLYT 1131

Qy        324 GLRFIIKKASN----SRNINNDN-------------IVREGDYIYL-------------- 352
              |    ||  |:    || :| ||             |:|: | ||
Db       1132 GNPITIKSVSDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYAL 1191

Qy        353 ----NIDNISDESYRVYVLVNSKEIQTQLFLAPINDDPTFYDVLQIKKYYEKTTYNCQIL 408
                  |: |      : :  :|:  :  :|:| :    :     |: :  ::  |  |    :
Db       1192 KLQSNLGNYGIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAV 1251

Qy        409 CEKDTKTFGLFGIGKFV-KDYGYV 431
                :||         ||: :| |:|
Db       1252 TNYETKLLSTSSFWKFISRDPGWV 1275


RESULT 15
Q841S3
ID   Q841S3      PRELIMINARY;      PRT;  1280 AA.
AC   Q841S3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotoxin.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW/2003;
RA   Lee M.S., Cheng M.C., Chen Y.P., Shiau J.R., Lin S.Y.;
RT   "An outbreak of botulism in black-faced spoonbill in Taiwan.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY251553; AAP06952.1; -.
DR   GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0015070; F:toxin activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0009405; P:pathogenesis; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR002160; Kunitz_legume.
DR   InterPro; IPR000395; Peptidase_M27.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1280 AA;  147642 MW;  34C1A2E573EE0AE7 CRC64;

  Query Match           23.8%;  Score 607;  DB 2;  Length 1280;
  Best Local Similarity   31.1%;  Pred. No. 6.4e-25;
  Matches  148;  Conservative   82;  Mismatches  168;  Indels   78;  Gaps   13;

Qy         27 DTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFNDIGNGQFKLNN 86
              :::|  : | | ::|: : |||| :   |:|:||| | : :  ||  |:|    |||::
Db        854 NSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRLEGDVQVNTIYTNDFKLSS 913

Qy         87 SENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTIISCIKNDSGWK 146
              |    |  : :  ::| ::::| |::||::  |   |       ||||||: || :||||
Db        914 S-GDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNS-----HNEYTIINSIKQNSGWK 967

Qy        147 VSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLGNANIYINGSL 206
              : |:   | | | ||| | ||: |:||   : : | |||| :|||: :|   :|||| |
Db        968 LCIRNGNIEWILQDINRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYINGEL 1027

Qy        207 KKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVSSLYWIQSSTN 266
              |:|:|| :|: :     | | :    | : :||:|||| :||:  ::: :|  |  |
Db       1028 KQSERIEDLNEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQILRN 1087
```

```
Qy        267 TLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAINY---QNLYL 323
              :||:|||||::||:||: |    : ||           ||::|     : |    -||
Db       1088 VIKDYWGNPLKFDTEYYIINDNYIDRYI----------APKSNI-LVLVQYPDRSKLYT 1135

Qy        324 GLRFIIKKASN----SRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQLFL 379
              |   ||  |:   . || :| |||:                   ::| || :
Db       1136 GNPITIKSVSDKNPYSRILNGDNIM-----------------FHMLYNSGKYMI---- 1172

Qy        380 APINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIG---------------- 422
               | |  | | :        : : || :   |     | :|||
Db       1173 --IRDTDTIYAI-----EGRECSKNC-VYALKSQSNLGNYGIGIFSIKNIVSQNKYCSQI 1224

Qy        423 --KFVKDYGYVWDTYDNY-FCISQWYLRRISENINKLRLGCN--WQFIPVDEGWTE 473
                |:|:   : | |  : |     |      |     |  :  |:||  | || |
Db       1225 FSSFMKNTMLLADIYKPWRFSFENAYTPVAVTNYETKLLSTSSFWKFISRDPGWVE 1280
```

Search completed: September 10, 2004, 03:41:54
Job time : 23.6421 secs